Butler, Jeffrey V.; Giuliano, Paola; Guiso, Luigi

Working Paper

Trust and cheating

IZA Discussion Papers, No. 6961

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Butler, Jeffrey V.; Giuliano, Paola; Guiso, Luigi (2012) : Trust and cheating, IZA Discussion Papers, No. 6961, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/67185

DISCUSSION PAPER SERIES

IZA DP No. 6961

Trust and Cheating

Jeffrey V. Butler
Paola Giuliano
Luigi Guiso

October 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Trust and Cheating

Jeffrey V. Butler
EIEF

Paola Giuliano
UCLA, NBER, CEPR and IZA

Luigi Guiso
EIEF and CEPR

Discussion Paper No. 6961
October 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 6961
October 2012

ABSTRACT

Trust and Cheating*

When we take a cab we may feel cheated if the driver takes an unnecessarily long route despite the lack of a contract or promise to take the shortest possible path. Is our decision to take the cab affected by our belief that we may end up feeling cheated? Is the behavior of the driver affected by his beliefs about what we consider cheating? We address these questions in the context of a trust game by asking participants directly about their notions of cheating. We find that: i) both parties to a trust exchange have implicit notions of what constitutes cheating even in a context without promises or messages; ii) these notions are not unique – the vast majority of senders would feel cheated by a negative return on their trust/investment, whereas a sizable minority defines cheating according to an equal split rule; iii) these implicit notions affect the behavior of both sides to the exchange in terms of whether to trust or cheat and to what extent. Finally, we show that individuals' notions of what constitutes cheating can be traced back to two classes of values instilled by parents: cooperative and competitive. The first class of values tends to soften the notion while the other tightens it.

JEL Classification: A1, A12, D1, O15, Z1

Keywords: trust, trustworthiness, social norms, culture, cheating

Corresponding author:

Paola Giuliano
Anderson School of Management
UCLA
110 Westwood Plaza
Entrepreneurs Hall C517
Los Angeles, CA 90095
USA
E-mail: paola.giuliano@anderson.ucla.edu

* We thank Martin Dufwenberg, Uri Gneezy and Roland Benabou as well as seminar participants at EIEF and conference participants at the Sciences Po/IZA Workshop on Trust, Civic Spirit and Economic Performance, the 2011 Florence Workshop on Behavioural and Experimental Economics and the 2011 SITE Summer workshop at Stanford University for many helpful comments.

1 Introduction

When taking a cab we may expect the driver to use a reasonably short route even if neither we nor the driver make explicit mention of it. Despite the lack of explicit contract or promise, we may still feel cheated or disappointed if the taxi driver takes an unnecessarily long route. Similarly, when we ask for financial advice the advisor does not typically spell out that he will act in our best interest, but we may still judge cheating according to this metric. When we book a vacation through a travel agent, search for the best medical insurance at a broker or take our cars to a mechanic, we act on an implicit notion of what the behavior of the travel agent, broker or mechanic *should be*, perhaps feeling cheated or let down when behavior fails to live up to these standards.

Situations like these come up frequently in our daily economic lives: opportunities for mutually beneficial exchanges where complete contracts or agreements about what is expected from each side of the exchange are either impossible or infeasible. Considering only our first example above, according to one source over $600,000$ taxi rides are taken daily in New York city alone constituting about $1 billion in fares paid per year. [1] And New York is not alone: about one *million* people use taxis every day in Hong Kong,[2] while a staggering three to four million taxi rides are taken every day in Lima, Peru (Castillo, et al., 2012). Our second example—financial advice from professionals—is also pervasive. According to a broad survey of retail investors in Germany, more than 80 percent of investors consult a financial advisor. Overall, in the UK 91% of intermediary mortgage sales are "with advice" (see Chater, Huck and Inderst (2010)). In the US, 73% of all retail investors consult a financial adviser before purchasing shares (Hung et al., (2008)).[3]

In light of their ubiquity, understanding precisely what drives behavior in such situations is an important undertaking. In this paper, we focus on one intuitively plausible yet under-explored determinant of behavior on both sides of such exchange opportunities: individuals' personal, subjective, notions of what constitutes cheating. We investigate the role of cheating notions in the context of a trust game (Berg, Dickhaut and McCabe, 1995), a two-player sequential moves game of perfect information. In this game, the sender moves first by deciding whether to send some, all or none of a fixed endowment to a co-player, the

[1] http://en.wikipedia.org/wiki/Transportation_in_New_York_City.
[2] http://www.gov.hk/en/about/abouthk/factsheets/docs/transport.pdf
[3] See also Inderst and Ottaviani (2012) for a general review on financial advice.

receiver. Any amount sent is increased by the experimenter before being allocated to the receiver, who then decides whether to return some, all or none of this (increased) amount to the sender. While highly stylized, the trust game is an appropriate context because it captures the essential nature of our motivating examples: a pareto-improving exchange is possible, but comes with the risk of opportunistic counterparty behavior which cannot be eliminated through pre-play promises or contracts.

The timing of our experiment is as follows: first, we have participants play a (slightly-modified) trust game; after playing the trust game, we ask participants directly about their implicit cheating notions;[4] finally, we elicit participants' beliefs about others' strategies and cheating notions as well as their (second-order) beliefs about what others' expect from them. We complement the data from our trust game experiment with data on the values our participants' parents emphasized during their upbringing. We use these data to investigate one potential source of cheating notions.[5]

In this specfic setting we test several hypotheses. At the most basic level, we test whether such situations do indeed engender implicit cheating notions. Whether or not this will be the case is not *a priori* obvious: cab drivers, mechanics and financial advisors may very well choose to ignore or downplay the possibility that their customers could ever feel cheated in order to reconcile opportunistic behavior with a positive self-image.[6] Secondly, conditional on an affirmative answer to our first question we test the hypothesis that these implicit cheating notions have an impact in determining behavior in a trust-exchange situation.

We find that the vast majority of participants articulate a cheating notion even when they can easily refrain from doing so, suggesting they are genuine. We document these notions, showing they are roughly bi-modal: many participants define cheating by a positive return on investment rule, as assumed but not tested by Berg, Dickhaut and McCabe (1995); while, contrary to the assumptions of much of the trust game literature, a sizable minority of senders (around 30% of participants) define cheating by a more demanding notion requiring fully half of their co-players' total earnings in order to not feel cheated.[7]

[4] We realize that asking about cheating notions directly gives rise to concerns about priming. We check for the robustness of directly-revealed cheating notions in additional robustness sessions where cheating notions are elicited indirectly in a way that reduces the likelihood of priming effects.

[5] The data were collected for a previously-conducted, unrelated, experiment.

[6] For evidence that individuals choose their beliefs to avoid cognitive dissonance, we refer the interested reader to the discussion in Akerlof and Dickens (1982).

[7] Because of the way we modified the trust game, this latter rule can be distinguished from previously documented fairness rules such as equal *surplus* division. For details, see the experimental design section.

On our second point, we find that the notion of cheating strongly affects behavior on both sides of the potential exchange. On the sender side, we find an inverse relationship between beliefs about the likelihood of being cheated and the amount sent. On the other side, we find that receivers' behavior varies significantly with their beliefs about senders' cheating notions. When studying receivers' behavior, we define "intentional cheating" as occuring when a receiver sends back strictly less than the receiver *him/herself* believes the sender needs back in order to not feel cheated. We find that the return amounts of receivers who do not intentionally cheat vary one-to-one with their beliefs about senders' cheating notions, while the return amounts of those who do intentionally cheat are consistently about half as sensitive to their beliefs about others' cheating notions.

Put simply, even receivers who choose to cheat do not go all the way to returning nothing; rather they typically return something. If they choose not to cheat then they give back the absolute minimum amount that allows them to avoid the psychic cost of cheating—i.e. the amount dictated by the expected cheating notion of the trusting person in the exchange.

Having shown the importance of implicit cheating notions for behavior, we go one step further and investigate their determinants. In our investigation of intentional cheating, we discovered that receivers with higher standards—i.e., those who would feel cheated unless they are given back a lot when playing as senders—were consistently less likely to cheat irrespective of send amount. This suggests that own cheating notions may be related to the intensity with which individuals care about good behavior. We test this intuition using responses from an unrelated survey asking participants about the values their parents instilled during their upbringing. We classify instilled values into two categories: cooperative and competitive. In the cooperative category we place values that might promote fair exchange, such as "always give others their fair share." The competitive category includes values that encourage individuals to compare themselves and their actions to others. An example of the latter category is "always seek to be better than others." Consistent with the idea that cheating notions proxy for concern for good behavior, we find that both types of values matters for individuals' cheating notions. Controlling for both categories of values simultaneously, we find that they pull in opposite directions: instilled cooperative values tend to decrease the amount of money one needs back in order to not feel cheated; while emphasis on competitive values tends to increase own cheating notions. Hence, differences across individuals in cheating notions are partly a reflection of the relative emphasis parents

place on these two classes of values.

The remainder of the paper proceeds as follows: Section 2 discusses closely related literature; Section 3 details the experimental design; Section 4 presents the results; Section 5 provides a more general discussion of our findings together with a simple analytical framework to interpret them. The final section concludes and suggests avenues for future research. Additional experimental treatments, analyses conducted to address the robustness of elicited cheating notions, and a comparison between behavior in our main experiment—conducted on-line—and a smaller study conducted in a more traditional laboratory environment can be found in Appendix I. Appendix II provides instructions for our main experiment.

2 Closely related literature

Our paper is closely related to the huge literature investigating behavior in the trust game.[8] The bulk of this vast literature focuses on trust game senders' behavior, interpreting the amount senders send as "trust," whence the label "the trust game" stems. What, precisely, senders are trusting receivers to do is typically left unspecified, but a common assumption—made explicitly in Berg, Dick and McCabe (1995)—is that senders are trusting that receivers will send back at least as much money as they sent. To the best of our knowledge, this assumption has never been tested directly. Our paper contributes to the trust game literature by shedding light on what it is, exactly, that senders are trusting receivers to do. Furthermore, by controlling for senders' beliefs about being cheated according to senders' own cheating notions, we provide novel evidence that the trust game actually involves trust which is a surprisingly contentious point.[9] The most closely related papers in this vein are Bohnet and Zeckhauser (2004) and Cox (2004), which both provide *indirect* evidence that senders' behavior in the trust game measures trust. Bohnet and Zeckhauser (2004) show that the extensive margin of trust (whether to send anything) depends on expected cheating or betrayal, while we extend this finding to show that the the intensive margin of trust (how much to send) varies with expected cheating. Cox (2004) decomposes the standard trust game into three closely related, but simpler, games and shows that sender behavior

[8] The trust game literature is too large and spans too many disciplines to be summarized here, but for an excellent review see Camerer (2003) and the references therein.

[9] Trust is, itself, notoriously difficult to define precisely. In making this statement, we rely on a broad definition stemming from an inter-disciplinary (including economics) review of the trust literature: trust is the sender's "...intention to accept vulnerability based upon positive expectations of the intentions or behavior of others..." (Rousseau, et. al., 1998).

in the trust game cannot be solely attributed to risk preferences or altruism. We extend this result by directly showing that, controlling for risk preferences, altruism and a host of demographics, senders' decisions are motivated by expectations of non-malign counterparty behavior and hence involve trust.

A second literature closely related to our inquiry focuses on what constrains opportunistic behavior of the entrusted in situations like the trust game. An influential literature models this behavior as being driven by an aversion to guilt arising from disappointing others (Dufwenberg and Gneezy, 2000; Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007). Disappointment is defined relative to (first order) beliefs about what others will do and, consequently, guilt is defined with respect to individuals' second-order beliefs about what others expect from them. This "guilt aversion" literature, however, leaves unspecified what determines first- and second-order beliefs about others' actions and expectations. We contribute to this literature in two ways: i) by providing evidence that, while conceptually distinct from first- and second-order beliefs, implicit cheating notions are in fact highly correlated with individuals' beliefs about what others will do and what others expect them to do;[10] and ii) by showing that an important source of cheating notions, and hence first- and second-order beliefs concerning others' actions, is parentally instilled values.

More generally, our results contribute to the debate over how non-pecuniary preferences affect behavior and where these preferences come from. Receivers in the trust game face a stark trade-off between their pecuniary preferences and moral behavior. Our finding that receivers' behavior is affected by their beliefs about what constitutes cheating lends support to the view put forward by Gneezy (2005): moral preferences are affected by the magnitude of damage that immorality inflicts on others.[11] However, because our experiment involves a game with neither communication nor unambiguous moral standards, and hence no literal lying nor deception, we extend Gneezy's findings by showing that the moral forces at work operate outside of the specific context of deception. We inquire why people act this way,

[10] Notice that if senders' expectations of receivers' behavior vary with senders' implicit notions of cheating—for whatever reason—this provides a channel for receivers' beliefs about senders' cheating notions to enter into receivers' (second order) beliefs about what senders expect from them.

[11] Many popular and intuitive models of moral preferences are inconsistent with this pattern in behavior. For an elaboration of the inconsistencies, see Gneezy (2005). As but the most obvious example, notice that fixed-cost-of-immorality models imply that increasing the benefit of immorality increases immorality irrespective of damage to others which is at odds with observed patterns in behavior in both our experiment and in Gneezy's experiments.

identifying values instilled by one's parents as an important reason.

3 Experimental Design

A total of 428 individuals participated in our main study, all of whom were students in Rome, Italy, enrolled at one of two universities: LUISS Guido Carli University or the University of Rome, La Sapienza. All sessions were conducted on-line.

This experiment consisted of three phases. First, participants played a slightly modified trust game. Responses were collected using the strategy method, so that each participant submitted decisions in both the sender and receiver role before knowing which role they would be assigned. Next, they were asked about their subjective cheating notions from the point of view of the sender role. Finally, we elicited participants' beliefs about others' behavior and others' cheating notions in the trust game. During each of these three phases, participants were unaware of the existence of any of the subsequent phases.

3.1 Our mostly-standard trust game

In our trust game specification, players are assigned one of two roles—sender or receiver. Each sender is endowed with 10.5 euros and randomly and anonymously paired with another participant—the receiver—who is given no endowment. The sender moves first and chooses whether to send some part of his or her endowment to the receiver, or to keep the entire endowment. Feasible send amounts consist of any integer between 0 and 10, inclusive. Any money sent is increased by the experimenters according to a concave function before being allocated to the receiver. If the sender sends S euros, the receiver receives (approximately) $8S^{0.5}$ euros.[12] The receiver then chooses how much, if any, of the received amount to send back to the sender and the game ends.

The main difference between our game and the standard trust game is our concave "trust production function." This modification serves two purposes. First, for low send amounts concavity permits stronger separation of plausible cheating notions than the more standard

[12] Since the sender can send only integer amounts, the possible amounts a receiver could receive are: $f(1) = 8.05, f(2) = 11.30, f(3) = 13.85, f(4) = 16.05, f(5) = 17.90, f(6) = 19.60, f(7) = 21.20, f(8) = 22.65, f(9) = 24.05, f(10) = 25.30$. This trust production function deviates slightly from $f(S) = 8S^{0.5}$ in order to produce relatively simple values (multiples of 5 cents) while, at the same time, maintaining concavity and surplus creation. Surplus creation is a central feature of the trust game and refers to the fact that each additional dollar sent always produces more than one dollar in receiver earnings. Participants were presented the trust production function in table format to facilitate comprehension.

linear production function, $f(S) = 3S$. For example, consider the first euro sent. This euro produces 3 euro in receiver earnings in the standard trust game, while it produces 8.05 euros in our implementation. In the standard trust game, as in ours, a cheating notion defined by a positive return on one's "investment" requires that receivers return (at least) 1 euro. Another possible cheating notion is that the sender requires half of all the money produced by the action of sending in order to not feel cheated. This latter notion requires receivers to return 1.50 (4.025) euros using the standard (our) trust production function. Consequently, these two notions would differ by only 0.50 euro using the standard trust production, while, in contrast, our concave trust production function separates these two cheating notions by about 3 euros (4.025 euros vs. 1 euro), making it less data-intensive to distinguish the former notion from the latter.

The second purpose of a concave production function is to provide a relatively smooth relationship between trust behavior and trust beliefs, aiding our examination of the "intensive margin" of trust (how much to send conditional on sending something). For instance, if senders have standard risk-neutral preferences a linear trust production function often implies corner solutions: send the entire endowment if the expected return from trusting is positive, or nothing if the expected return is negative; if the expected return is zero, then all send amounts are optimal. In contrast, our concave production function provides such senders unique internal optimal send amounts that vary continuously with monetary return beliefs for a wide range of beliefs.

Participants' trust game strategies were collected using the strategy method. Before discovering whether they would play the role of sender or receiver, participants submitted a complete contingent strategy for each role: each participant specified how much they would send in the role of sender; for each possible amount they could receive in the role of receiver, each participant specified how much they would return. The order in which participants submitted their strategies—whether first for sender, then for receiver or first for receiver and then for sender—was randomized. Additionally, to bridge the gap between the strategy method and the direct response method and to attempt to make each receiver's decision feel as real as possible, participants' receiver strategies were elicited with a series of ten separate screens. Each of these ten screens asked only one question: "if the sender sends S euros and you therefore receive $f(S)$ euros, how much will you return?"[13]

[13] For each separate screen, S was replaced with exactly one of the 10 possible amounts a sender could send

To help our analysis of the intensive margin of trust and to separate it from the determinants of the decision to trust at all (the extensive margin of trust), in some sessions we imposed a "sending fee" of 0.50 euro. In these sessions, senders who chose to send a strictly positive amount incurred the fee whereas senders who chose to send nothing were not charged the fee. We call sessions involving this sending fee "high fee" sessions. We label the remaining sessions—where no sending fee was charged—"low fee" sessions. In both high fee and low fee sessions, senders had the same set of feasible actions: $S \in \{0, 1, \ldots, 10\}$.

3.2 The cheating notion questions

After participants submitted their complete contingent strategies, we asked them to specify their personal definitions of cheating from the perspective of the sender. For each possible strictly positive send amount, $1 \leq S \leq 10$, participants were asked:[14]

> "If you are assigned the role of A [sender] what is the minimum amount you would need to receive back from player B [receiver] in order to not feel cheated? ... If you were to send S euros and B were to therefore receive $f(S)$ euros, you would need back how many euros?"

To respond, participants could either insert a number between 0 and $f(S)$ or refrain from specifying a cheating notion by selecting one of two options: "this has nothing to do with cheating;" or "I don't know." Leaving the question blank was also possible, but not explicitly mentioned as an option.

Initially, our design did not include these two explicit opt-out responses. Although responding to the question was always completely voluntary, we realized that not providing pre-programmed opt-out responses could make some participants feel obligated to supply a cheating notion even if they did not truly have one. To address this concern, we inserted the two opt-out responses. The majority of participants—306 out of 428—took part in sessions featuring the explicit opt-out responses. The remaining 122 participants took part in sessions with no explicit opt-out opportunity. Unless otherwise specified, our analyses

($S \in \{1, \ldots, 10\}$), and $f(S)$ was replaced with the corresponding value from the trust production function ($f(S) \in \{8.05, \ldots, 25.30\}$). The order in which receivers faced their ten separate decisions was randomly predetermined but the same for all participants. This maintains comparability across observations without inducing undue consistency in receiver strategies that might arise from, e.g., facing a monotonically increasing or decreasing sequence of send amounts. The order used was $S = 7, 4, 8, 3, 9, 10, 2, 1, 6, 5$.

[14] In each question "S" and "$f(S)$" were replaced by the appropriate numbers. The words "sender" and "receiver" did not appear on participants' screens.

utilize all 428 observations. In Appendix I we show that our results are robust to restricting attention only to sessions with explicit opt-out.

One additional concern with our direct cheating notion question is priming. Some may argue that by asking about cheating directly we may be priming participants to associate behavior in the trust game with cheating. To address this concern we ran additional sessions in which, rather than asking our direct cheating notion question above, we asked participants to state how they would feel about various send/return combinations if they were to be assigned the role of sender. The results support the idea that priming is not the driver of reported cheating notions. Participants in fact single out negatively valenced adjectives ("cheated" or "disappointed") to describe their feelings even when there are other options available. For further details of these sessions and results, see Appendix I, section A.2.

3.3 The beliefs elicitation phase

Following the cheating notions questions, participants discovered there would be a beliefs elicitation portion of the experiment, and that they could earn additional money according to the accuracy of their estimates. Beliefs were paid according to a randomized quadratic scoring rule (Schlag and van der Weele, 2009) which is both incentive compatible and theoretically robust to risk preferences. Participants were informed that one estimate from this section would be chosen to count toward their potential earnings. An exactly correct belief paid 5 euros in most sessions. In two sessions we strengthened belief elicitation incentives—exactly correct beliefs paid 20 euros—in order to allow us to investigate the importance of ex-post rationalization (Appendix I, section B).

In the beliefs elicitation section each participant was asked to estimate: i) how much other senders would send on average; ii) how much other receivers would return on average; iii) their beliefs about others' beliefs about how much receivers would return (second order beliefs); iv) other participants' cheating notions; and v) the proportion of other participants who would not cheat them, according to the respondent's own subjective cheating notion (see Appendix II for exact wording).[15] For all belief elicitation questions, participants were instructed to exclude their own actions from their estimates and were told that the accuracy of their estimates would be calculated excluding their own strategies and cheating notions. This was done to avoid mechanical correlations between reported beliefs and participants'

[15] Items ii)-v) were asked for each possible send amount.

own strategies or cheating notions. Beliefs were elicited *after* participants submitted their complete contingent strategies, but *before* knowing their assigned roles.

3.4 Payment phase

After all three phases of the experiment were completed, pairings were randomly determined and within each pair roles were randomly assigned. Outcomes and potential earnings were determined by combining, within each pair, the sender's strategy with the receiver's strategy. Participants were informed at this point which randomly-chosen belief elicitation question would count toward their potential earnings and how much their estimates earned them. Finally, 10% of participant pairs were randomly chosen to be paid their potential earnings.

While 10 percent may seem low, the experiment was relatively short and convenient, requiring on average about half an hour of participants' time. Furthermore, note that Italian students' opportunity costs are relatively low. As an example, work-study positions at one university in Rome we are familiar with typically pay students around 5 euros per hour. Given both of these observations, we feel the expected earnings from the experiment are commensurate with participants' opportunity cost of time. Finally, note that behavior in our in-lab sessions, where all participants were paid, was remarkably similar (see Appendix I, section A.1), suggesting that the relatively weak incentives in our on-line study were nevertheless adequate. The design of the experiment is summarized in Table 1.

3.5 Instilled values and risk attitudes

For each participant in our main study, we complement the experimental data with data from a previously conducted, unrelated, survey. This survey contains basic demographic information, a (self-reported) measure of the emphasis each participant's parents placed on various normative values during his or her upbringing as well as an incentive-compatible measure of risk aversion (Holt and Laury, 2002).[16]

[16] Briefly, this procedure asks participants to make a sequence of ten choices, each of which involves a choice between a relatively risky lottery (38.50 euros with probability p, 1 euro with probability (1-p)) and a safer lottery (20 euros with probability p, 16 euros with probability (1-p)). The probability of the high payoff, p, increases over the sequence from 0.1 to 1.0 in steps of 0.1. This construction implies that more risk averse individuals will switch from preferring the safer lottery to the riskier lottery later in the sequence. We use the choice number in the sequence where this switch occurs as our risk preferences measure which ranges from 1 to 10 and is increasing in risk aversion. For ten percent of survey participants one decision in this sequence was randomly chosen and these participants were paid according to the outcome in their preferred lottery.

There was a considerable time lag between the survey and the start of our trust game experiments (from 20 to 60 days) so that survey responses are unlikely to have affected trust game behavior directly. On the other hand, this temporal distance was small enough so that traits, such as risk aversion or instilled values, likely did not change in the meantime. This survey data allows us to control for risk aversion and altruism when examining sender behavior, while instilled values will prove useful in examining what drives receiver behavior.

4 Results

We establish four main results: *i)* the vast majority of trust game participants have a cheating notion; *ii)* beliefs about cheating are relevant in senders' decisions to trust; and *iii)* controlling for receivers' expectations about senders' cheating notions, receivers that have higher standards—i.e., would feel cheated unless they were given back a lot when playing as senders—are consistently less likely to cheat across all send amounts; moreover, *iv)* intergenerationally transmitted values are important determinants of receivers' cheating notions. Descriptive sample statistics are reported in Table 2.

4.1 Notions of Cheating

We start by remarking that the vast majority of senders—about 80%, depending on send amount—have a notion of cheating even when they are given the possibility of not reporting one. Restricting attention to sessions involving explicit cheating notion opt-out, the proportion of senders selecting the option "this has nothing to do with cheating" ranges from a low of 13 percent when considering sending 10 euros, to a high of 20 percent when considering sending one euro.[17] These proportions are summarized in Table 3.

Next, we examine the cheating notions that senders report. Table 4 reveals that, conditional on reporting a cheating notion, most senders report one that is *at least as demanding* as a positive return on investment rule. That is to say, most would feel cheated if their co-player fails to return at least as much as is sent and many would feel cheated by an even larger return amount. This is true for any amount a sender could possibly send, and irrespective of whether a sender could opt out of reporting a cheating notion. Considering all sessions pooled, the fraction of senders who would feel cheated by a negative return on

[17] The proportion of senders who opt out of reporting a cheating notion for *any* reason—which includes selecting "I don't know" or just leaving the question blank, as well as selecting "this has nothing to do with cheating"—in these same sessions is also low, ranging from 17 percent to 23 percent (Table 3).

their investment (conditional on specifying a cheating notion) ranges from 91 percent down to 79 percent with higher fractions for smaller amounts sent.[18,19]

As a second step we plot kernel density estimates for individuals' own cheating definitions and their estimates of others' cheating notions. We begin by considering cheating notions conditional on sending 1 euro (Figure 1). Participants' own cheating notions and their estimates of others' cheating notions are similarly distributed and roughly cluster around 1 euro and 4.025 euros (indicated by the two vertical bars). The first value is consistent with cheating being defined according to a positive return on investment rule, while the latter value suggests cheating defined with respect to an equal split of the entire amount receivers receive—about 8 euros in this case.[20] While the latter cheating notion may seem extreme, it fits with a common interpretation of trust game receivers' situation: they are essentially in the position of dividing a fixed-size pie and, in such situations, a common notion of fairness dictates dividing the sum into equal shares (see, e.g., Camerer, 2003). However, notice that since we use an unequal endowment design, equally splitting the receiver's income implements a very unequal outcome: about 13 euros for the sender and around 4 euros for the receiver.

Figures 2A-2C present the analogous kernel density plots for the rest of the possible send amounts, first for sessions without explicit opt-out opportunities (Figure 2A), then for sessions featuring explicit opt-out choices (Figure 2B) and, finally, for all sessions pooled (Figure 2C). There are two points to notice. First, regardless of whether explicit opt-out responses were available, own cheating definitions follow the same patterns as in Figure 1: own cheating notions are bi-modal and concentrated around an equal-split rule and a positive return on investment rule. As these two definitions become closer to one another (which happens at larger send amounts because of the concavity of our trust production function) own cheating notions become more nearly unimodal. Secondly, also beliefs about

[18] Differences in proportions tests fail to reject equality of these percentages across explicit-opt-out and no-explicit-opt-out sessions, for each send amount separately, at the 5 percent level.

[19] In Table 4 the condition that defines cheating is *amount returned - amount sent*< 0. It ignores the 50 cent fee and in this it is consistent with the wording of the question that we asked to elicit cheating notions. An alternative would be to define cheating as *amount returned - (amount sent*$+0.50) < 0$ in sessions with a strictly positive investment fee. Results when this alternative criteria is used are similar but all proportions shown in Table 4 are slightly lower.

[20] Using Hartigan's test for unimodality, the null hypothesis that the empirical distribution of own cheating notions when sending 1 euro is unimodal can be rejected ($p < 0.01$). This is true also if one restricts attention to only those sessions where refraining from specifying a cheating notion was possible which is less obviously bi-modal. Unimodality can also be rejected for the distributions of estimated others' cheating notions when sending 1 euro ($p < 0.01$).

others' cheating notions continue to roughly cluster around the two norms of equal-splits and positive return on investment across all send amounts. All together these patterns demonstrate that even in sessions where participants gravitate toward an equal-split notion of cheating at the expense of a positive return on investment rule, they believe that others will cluster around both of these notions of cheating.

In terms of reported beliefs, we find that first order beliefs about others' cheating notions are highly correlated with receivers' second order beliefs about what they believe senders expect back from them, suggesting that receivers believe senders expect receivers will not cheat them. The raw correlation between these two variables ranges from a high of 0.66 conditional on sending 1 euro to a low of 0.58 when considering a send amount of 8 euro. In all cases the correlation is highly significant ($p < 0.001$).

Our analysis of the kernel densities suggests that implied notions of cheating fall roughly into two categories. While a positive return on investment rule serves well as a conservative estimate of what most participants would consider cheating, using only this notion mischaracterizes a large proportion of participants. A sizable minority—roughly one third of participants—demand *half of the total money receivers receive* in order to not feel cheated. We label these latter participants "equal splitters." As is evident in Table 5, the overall proportion of equal splitters for each send amount is fairly constant and always around one-third.[21] To see whether individuals are consistent with respect to their cheating notions, we restrict attention to participants who are equal-splitters conditional on sending 1 euro and plot the kernel density estimates of these individuals' cheating notions across other send amounts. As is evident in Figure 3A, equal splitters are consistent: those who define cheating in terms of an equal-split when sending 1 euro tend to use this cheating definition for other possible send amounts. These same individuals also tend to believe others define cheating in the same way consistently across send amounts (Figure 3B).

To sum up, the vast majority of people playing the trust game have clear notions of cheating even though receivers make no explicit promises. Having established this, the obvious question to ask is whether expected cheating drives senders' behavior. We now

[21] Because experimental participants have a well-known proclivity to state integer values when possible, we include in our definition of equal splitters any subject who stated a cheating notion between the largest integer less than, and the smallest integer greater than, half of the total amount receivers receive in order to not feel cheated. For example, for a send amount of 1 euro, receivers receive 8.05 euros, and half of this amount is 4.025 euros. Consequently, our definition of an equal-splitter conditional on sending 1 euro includes anybody who reported a cheating notion between 4 and 5 euros.

turn to addressing this question.

4.2 The Effects of Cheating Beliefs on Senders' Behavioral Trust

One of our main results is that participants' trusting behavior depends crucially on how likely they think it is that they will be cheated, where cheating is defined according to trustors' own personal notions of being cheated. To show this, for each participant we construct a unidimensional measure of his or her beliefs about the proportion of non-cheaters in the (experimental) population. We do this by averaging each participant's answers to the following set of 10 questions ($S = 1, 2, \ldots, 10$):

> "If you send S euros and B therefore receives $f(S)$ euros, what percent of B's will return enough money so that you do not feel cheated?"

The resulting measure of beliefs about population trustworthiness theoretically ranges from 0 to 1, with 1 indicating the sender believes no receiver will cheat for any send amount (all are trustworthy) and 0 indicating all receivers will cheat for every send amount (none is trustworthy).

Before proceeding we must address one technical issue: how to construct this measure for individuals who, for a particular amount sent, report no cheating notion. First of all, if an individual responds that sending S euros "...has nothing to do with cheating," then it is reasonable to assume that this individual *cannot* feel cheated regardless of the receiver's decision. Therefore, we code such individuals' population trustworthiness belief conditional on sending S euros as 1 before constructing the summary measure.[22] On the other hand, if an individual did not report a cheating notion conditional on sending S euros for any other reason, then our elicitation mechanism is not incentive compatible since we cannot observe whether such an individual will feel cheated. For these participants, we code as missing their answer to the population trustworthiness question associated with sending S euros, which also results in a missing observation with respect to our summary measure.

Given these caveats, we construct a unidimensional measure of population trustworthiness for 401 (out of 428) participants, which we interpret as their subjective probabilities

[22] This could be problematic for our analysis if the subset of people who consistently report that sending money has nothing to do with being cheated also sends more on average. However, this does not seem to be the case. Only 39 individuals have a population trustworthiness measure equal to 1. The average send amount for these 39 individuals is 4.28, which is not significantly different from the average send amount for the remaining 362 individuals (4.34).

of not being cheated. Figure 4 plots the kernel density of this probability separately for opt-out and no-opt-out sessions. We document a modal value at around 0.5 (almost equal to the fraction of non cheaters in the pool—see Table 2, bottom row) irrespective of opt-out opportunities. In sessions with opt-out (the dashed line), a second mass of observations centers around a value of 1, reflecting (mechanically) the small minority of participants who report the trust game "has nothing to do with cheating" consistently.

In an analogous fashion, we construct for each participant a summary measure of his or her beliefs about the proportion of the money they send that will be returned to them. For each $S, 1 \leq S \leq 10$ we divide the participant's estimated return *amount* conditional on sending S euros by S to get their estimated (gross) return proportion conditional on sending S. We then average their 10 return proportion estimates to get a unidimensional measure of return proportion beliefs. The resulting averages range from a low of 0.00 to a high of 4.02 with a mean of 1.27 and a standard deviation of 0.64. We interpret this index as a measure of senders' expected (gross) return proportion and note that, on average, expectations are nearly identical to actual gross return proportions (Table 2).

Finally, using these two summary measures we estimate a model of how much senders send as a function of the senders' expected return proportion, their beliefs about being cheated and an interaction between these two variables. We control for a host of demographic variables. To account for selection into sending a positive amount we estimate a Heckman model and exploit variation in the investment fee across sessions to construct the selection equation. Specifically, the exclusion restriction for the selection equation consists of a dummy for "Low fee" sessions—i.e., sessions where the investment fee was zero. Importantly, because two common alternative explanations for senders' behavior in the trust game are risk preferences and altruism, among our demographic controls we include an incentive-compatible measure of risk aversion collected from the survey described in Section 2.5 as well as a proxy for altruism obtained from that same survey.[23]

Table 6 presents the estimates. Here and in all subsequent regressions we cluster standard errors by session. The estimates imply that the probability of being cheated plays a significant role in the intensive margin of trust: the positive and significant coefficient on our measure of the expected probability of *not* being cheated indicates that when senders

[23] We use as our measure of altruism the emphasis, on a scale from 0 to 10, participants' parents placed on the value of "helping others."during their upbringing.

believe it is less likely that they will be cheated, they send more. The implied effect of non-cheating beliefs on behavioral trust is large: increasing this belief from 0.1 to 0.9 is associated with an increase in the average amount sent equal to 51% of the sample mean. The coefficient on expected pecuniary returns is also positive and significant, indicating that standard pecuniary concerns also drive senders' behavior. Finally, the negative and (marginally) significant coefficient on the interaction between expected returns and non-cheating beliefs suggests that as expected pecuniary returns increase, the negative impact on trust of expected cheating subsides. In other words, the sting of expected betrayal can be soothed by money.[24]

4.3 What Drives Receivers' Decision to Cheat?

If expected cheating drives trusting behavior, then the question of what drives cheating becomes important for all the same reasons that trust itself is important. We therefore study what drives receivers' decisions to *intentionally* cheat. We can address this latter question directly because we know when receivers cheat according to their own estimates of others' cheating definitions.

We construct a dummy variable taking the value of 1 whenever receivers return less than they themselves believe their co-players need back in order to not feel cheated and 0 otherwise, for each amount a sender could send. This dummy indicates when receivers intentionally cheat. We then relate this variable to receivers' demographic characteristics as well as their own, and their estimates of others', cheating notions.

With regard to own cheating notions, we run into the same technical problem as before: how to handle participants who refrain from supplying a cheating notion. One potential answer is to code the cheating threshold as 0 whenever a participant responds "this has nothing to do with cheating," and as missing if they fail to supply a cheating notion for any other reason. This is what we do.[25]

Table 7 presents our estimates of receivers' propensities to intentionally cheat for each possible send amount. Participants' demographics have few consistent effects on cheating

[24] The results are virtually the same if we estimate a Tobit model of send amounts, which intuitively models selection as censoring.

[25] An alternate strategy of coding the cheating notion as missing whenever participants fail to report a cheating threshold for *any* reason roughly doubles the magnitude and increases the significance of all own cheating notion coefficients in Table 7. The same happens simply including a dummy for those who report that sending s euros has "nothing to do with cheating." The results presented should therefore be seen as a conservative estimate of the impact of own cheating notions.

across different send amounts: older participants generally cheat less for lower send amounts; smarter participants—those who have higher math scores—are less likely to cheat for high send amounts. Interestingly, gender plays no role. On the other hand, controlling for receivers' expectations about senders' cheating notions, receivers that have higher standards—i.e., would feel cheated unless they were given back a lot when playing as senders—are consistently less likely to cheat across all send amounts. We interpret this finding as saying that more demanding people tend to refrain from cheating others, behaving according to the principle "do not do to others what you would not want others to do to you." Notice, however, that conforming to this principle is cheaper when amounts sent are low and the temptation to deviate from it (and doing to others what you would not want them do to you) is thus weaker. Consistent with this we find that the effect of one's own cheating notion is stronger at low levels of amount sent and weaker at high levels: the reported probit coefficients imply that the marginal effect of an increase in receivers' own notions of cheating at send amount 10 is half that at send amount 1 (1.6 percentage points vs. 3.6 percentage points, respectively).

4.4 What Determines Cheating Notions?

Our finding that higher standards in own cheating notions reduce cheating suggests that cheating notions may be related to participants' moral values. Since these values tend to be culturally transmitted from parents to children (see e.g. Bisin and Verdier, 2010), we test whether participants' cheating notions are affected by the values their parents emphasized during their upbringing—instilled values—while controlling for our standard set of demographic variables.

We use data from a previously conducted unrelated survey (described in Section 2.5) which included a section on parentally instilled values. The survey asked about a rich set of normative values. For each normative value in this set, survey participants were asked to state how much emphasis their parents placed on this value during their upbringing. Valid responses ranged from 0, which indicates no emphasis, to 10 which indicated quite a lot of emphasis.[26] For our estimates, we select a relevant subset of these normative values and organize them into two categories: "cooperative" and "competitive." The former category includes such values as helping others and honesty. The latter category includes, for

[26] Participants could also respond "I don't know," which we code as missing.

example, the value of striving to be better than others.[27] We construct an index of parents' emphasis on "cooperative" and "competitive" values by taking the average emphasis over all the values constituting each category. This yields a measure for each category theoretically ranging from 0 and 10. We divide each of these measures by 10 obtaining an index on a 0 to 1 scale.

To get a summary measure of the impact of instilled values on cheating notions, we pool over all send amounts and regress cheating notions on cooperative and competitive values. Since pooling in this manner results in multiple observations for each participant we incorporate individual-level random effects in our model. We include a dummy variable for sessions with no explicit cheating notion opt-out. To control for the idea that the investment fee may directly affect cheating thresholds we also include a dummy for sessions with no investment fee. Finally, because our trust production function is approximately quadratic in money sent, we allow cheating notions to be a quadratic function of money sent.

Interestingly, estimates reported in Table 8 reveal that values matter for individuals' cheating notions but the two classes of values pull in opposite directions. Instilled cooperative values significantly lower cheating notions: the more emphasis parents placed on cooperative values, the fewer euros senders need back in order to not feel cheated. Competitive values, on the other hand, have the opposite effect, raising cheating notions significantly. Controlling for instilled values, cheating notions tend to move one-for-one with the amount sent, suggesting that a positive return on investment rule is the baseline cheating notion and that values determine how far individuals deviate from this baseline. Finally, there is little evidence that cheating notions vary by demographics, in particular gender, once we control for values.

5 Discussion and interpretation

In this section we try to put our results in perspective and shed light on the type of preferences that could explain receivers' cheating decisions. We start by plotting (Figure 5) the fraction of receivers who cheat at each send amount after partialling out the effect of the expected notion of cheating, thus purging the data from the mechanical effect this has on

[27] The full set of "cooperative" values is: i) behave as a model citizen; ii) help others; iii) group loyalty; iv) always give others their fair share; v) always tell the truth; vi) always keep your word. "Competitive" values are: i) always extract the maximum advantage from every situation; ii) seek to be better than others; iii) act so as to induce good in others (e.g., scold somebody who litters).

the probability of cheating. The share is 38% at send amount 1 and decreases, roughly, monotonically down to 1% at send amount 10.

This declining propensity to cheat as receivers receive larger sums from senders is inconsistent with both purely selfish preferences and fixed-cost of cheating models. With purely selfish preferences receivers would always cheat. On the other hand, since potential pecuniary gains from cheating increase in the amount sent, fixed cost of cheating models would predict a *non-decreasing* relationship between amount sent and cheating propensity.[28]

Patterns in our data also appear to be inconsistent with literal interpretations of many influential social preferences models. For example, consider inequality aversion (Fehr and Schmidt, 1999) or social welfare preferences (Charness and Rabin, 2002). Inequality averse individuals lose utility from unequal outcomes, while individuals with social welfare preferences place weight in their utility calculations on the outcome of the worst-off individual in their reference group as well as the total amount of money being distributed. In a two-player decision-making setting with no surplus creation opportunity such as that faced by trust game receivers, both of these models predict that receivers should never willingly put themselves behind in terms of final monetary payoffs.[29] However, a large fraction of receivers in our study do exactly that. For example, 82% of receivers willingly put themselves further behind than necessary when sent 1 euro and 47% of the receivers put themselves behind when sent 4 euros.[30]

The estimation results in Table 7 and the cheating pattern in Figure 5 could be justified in (at least) two ways. The standard justification is positive or negative reciprocity: sending more is a nicer action and/or sending less is a meaner action, so reciprocity demands responding in kind with a nice action (not cheating) or a mean action (cheating). An alternative explanation comes directly from the definition of trust: trust entails *vulnerability*.

[28] Let $B(S)$ denote the pecuniary benefit to the receiver from cheating, which increases with S, and assume the receiver j's fixed cost of cheating, K_j, is randomly drawn from the distribution $F(K)$. Then as S (and $B(S)$) increases, there should be a (weakly) higher proportion of receivers for which $B(S)$ exceeds K_j, and who therefore cheat.

[29] It is, of course, true that this strong prediction fails to hold if the receiver's reference group is something other than *just* himself or herself and his or her co-player. What the reference group is, or should be, is an important open question outside of the scope of this paper. We follow most of the literature in assuming participants view the trust game as a two-person interaction.

[30] This ignores the extra 50 cents senders have in sessions with no investment fee. Taking into account this extra 50 cents could obviously only increase these percentages. It should also be noted that receivers need not put themselves behind, except when the amount sent is 1. But even there, the figure of 82% includes only those receivers returning a *strictly* positive amount, so these receivers are willingly putting themselves further behind their counterpart than necessary.

At the same time, a widespread and intuitive moral standard is that, irrespective of what constitutes cheating, it is *more* wrong to cheat the more vulnerable. For example, cheating the elderly or the very young is commonly viewed as particularly reprehensible. This is the point made by Gneezy (2005). In the context of the trust game, sending more makes senders more vulnerable. Consequently, it is reasonable to assume that the moral costs of cheating increase in amount sent.

Without additional information we can rule out neither reciprocity nor vulnerability as the driving motive. To shed some light on why cheating declines in amount sent, in one of our robustness treatments we asked participants to describe their rationale for how they played the role of receiver in the trust game (Appendix I, section A.2.1). They could select one response from among four pre-programmed options and one free description. Three of the pre-programmed options were meant to capture positive reciprocity, negative reciprocity and vulnerability motives, respectively, while the fourth was essentially a "decline to state." option. We found that the modal response—selected by 42 percent of participants (72 out of 170) was the vulnerability explanation. The second most common response was positive reciprocity (about 30 percent of participants), while almost nobody chose negative reciprocity (6 percent; 10 out of 170 participants).

In light of these patterns, a unified way to model both senders' and receivers' preferences that is consistent with our data is to augment standard pecuniary preferences with a moral cost function. Individuals incur disutility from immoral actions, either when they are the perpetrator or the victim of such actions. Receivers lose utility when they cheat. Senders lose utility when receivers cheat them, which is consistent with our finding that expected cheating has a direct negative impact on the amount senders send. Beyond implying disutility from being cheated, our data do not say much about what senders' moral cost function might look like. On the receiver side, however, our data provide a bit more bite. For the rest of this section, therefore, we focus on receivers' preferences.

As a flexible specification for receivers' moral cost function, we assume it has three arguments: the vulnerability of the sender as measured by the amount sent, s; a fixed cost of cheating term; and a term measuring the degree with which the receiver cheats, as defined by the distance between the receiver's estimate of the sender's cheating notion and the amount the receiver returns, r.

In general, denote this moral cost function $m(s, K_j, dist(r, c_j(s)))$. A receiver's utility

is then given by:

$$U_j(r, s, c_j, K_j) = u(f(s) - r)) - \mathbb{I}(r < c_j(s)) \times m(s, K_j, dist(r, c_j(s))) \quad (1)$$

In (1), $f(s)$ denotes how much the receiver receives when the sender sends s and $\mathbb{I}(r < c_j(s))$ is an indicator function taking the value of 1 whenever the receiver intentionally cheats by returning less than dictated by the receiver's own estimate of the sender's cheating notion, $c_j(s)$. We assume that m is increasing in s, the vulnerability of the sender. We also assume that the fixed cost of cheating, $K_j \geq 0$, is a random draw from a common non-degenerate distribution function, $F(K)$. Finally, we assume that m is increasing and convex in its last argument, $dist(r, c_j(s))$, so that higher degrees of cheating are increasingly morally costly.

To be more concrete, a simple utility specification satisfying these assumptions is given by:

$$U_j(r, s, c_j, K_j) = u(f(s) - r)) - \alpha_j \mathbb{I}(r < c_j(s)) \times \{K_j + v(s) + \gamma_j (c_j(s) - r)^2\} \quad (2)$$

In equation 2, we assume that $u(f(s) - r)$, the receiver's standard pecuniary utility, is increasing and concave. The rest of the utility function captures the moral cost of cheating. The parameter α_j captures how much receiver j cares about morality. The parameter γ_j captures how much the receiver cares about degrees of cheating. A sender's vulnerability or niceness is captured by $v(s)$ which we assume is increasing.

There are three points to notice about this utility specification. First of all, setting $\alpha_j = 0$ reduces receivers' utility to standard (amoral) preferences. Secondly, notice that whenever $\alpha_j > 0$, setting $\gamma_j = v(s) \equiv 0$ implies receivers have simple fixed-cost-of-cheating preferences. Finally, if we assume that receivers expect senders to expect not to be cheated, then receivers' beliefs about senders' cheating notions—$c_j(s)$ in our model—may be closely linked to receivers' (second order) beliefs about how much money senders' expect receivers to return. In this case, our model can be thought of as an alternative way to capture guilt aversion which does not require knowledge of second order beliefs. Supporting this view, as already noted, in our data receivers' second-order beliefs are highly significantly correlated with their beliefs about others' cheating notions.

The specification for receiver utility can explain: a) why the decision to cheat depends on others' expected cheating notions; b) why cheating depends on the intensity of moral

preferences as proxied for by receivers' own cheating notions; and c) why the probability of cheating decreases in amounts sent as shown in Figure 5. This latter feature would be implied, for instance, whenever there are sufficiently many receivers with $\alpha_j > 0$ and when $v(s)$ is sufficiently steep in s. Intuitively, as $v(s)$ becomes steeper, cheating more vulnerable senders requires a larger offsetting pecuniary utility gain.

This simple preference specification can also account for another feature of the data: conditional on cheating, receivers on average do not go so far as to return nothing. Instead, they send something back. In our model, the amount returned by cheaters should depend positively on expected cheating notions, but—and this is the key prediction—it should *not* move one-to-one with the expected notion of cheating. On the other hand, conditional on not cheating, receivers should return the minimum amount consistent with satisfying the sender's notion. Non-cheaters' return amounts should therefore move one-to-one with the expected cheating notion.[31] Only the latter prediction is shared by both our model and the fixed cost of cheating model.

Table 9 puts these predictions to a test. We split the sample between cheaters and non-cheaters and estimate the amount receivers return as a function of their beliefs about

[31] The receiver's optimal choice can be found as follows. Suppose the receiver decides to cheat so that his or her utility is

$$U_j(r, s, c_j, K_j) = u(f(s) - r) - \alpha_j \times \{K_j + v(s) + \gamma_j(c_j(s) - r)^2\} \tag{3}$$

The amount the receiver sends back, r^*, is given by:

$$u'(f(s) - r^*) = 2\alpha_j\gamma_j(c_j(s) - r^*) \tag{4}$$

and is increasing in the estimated cheating notion $c_j(s)$ with a slope that is less than 1.
If the receiver decides not to cheat, the utility obtained is

$$u(f(s) - r) \tag{5}$$

$$subject\,to:\ r \quad \geqq \quad c_j(s) \tag{6}$$

and is maximized by setting $r = c_j(s)$ so that when the receiver does not cheat, the amount returned varies one-to-one with the expected cheating notion.

Finally, the receiver decides whether or not to cheat by comparing utility under the two cases and thus cheats if

$$u(f(s) - r^*(c_j(s))) - \alpha_j \times \{K_j + v(s) + \gamma_j(c_j(s) - r^*(c_j(s))^2\} > u(f(s) - c_j(s)) \tag{7}$$

or

$$u(f(s) - r^*(c_j(s))) - u(f(s) - c_j(s)) > \alpha_j \times \{K_j + v(s) + \gamma_j(c_j(s) - r^*(c_j(s))^2\} \tag{8}$$

where the left hand side is the net utility gain from cheating and the right hand side is the moral cost of cheating. This expression makes it clear that as s increases, provided $v(s)$ is sufficiently steep, cheating will diminish.

senders' cheating notions. We control for our standard set of demographics and, moreover, account for selection into cheating or not cheating by estimating a Heckman model. As the exclusion restriction in the selection equation, we use participants' own cheating notions. Broadly speaking the results are consistent with our model's predictions: the amount returned for both cheaters and non-cheaters depends on the expected notion of cheating, but return amounts are much more sensitive to expected cheating notions for non-cheaters than for cheaters. For non-cheaters, return amounts move essentially one-for-one with estimated others' cheating notions, while for cheaters this coefficient is consistently around half as large.

6 Concluding Remarks

Many real life exchanges require the "trustor" to decide whether and how much to trust a "trustee" who makes no promise on how he will behave in response to the trust received. This paper investigates whether individuals' personal, subjective notions of what constitutes cheating can drive this type of decision. We do so in the context of a trust game, where we elicit participants' definitions of being cheated as well as their beliefs about others' cheating definitions. We show that: i) participants have cheating definitions when playing the trust game, and that ii) these implicit notions affect the behavior of both sides to the exchange of whether to trust and by how much and whether to cheat and by how much; in addition, iii) trustors' and trustees' beliefs about what constitutes proper behavior need not to coincide. Our data reveal that the trust game gives rise to (at least) two cheating notions —a positive return on investment rule and an equal split rule.

After showing the relevance of implicit cheating notion, we investigate directly the determinants of untrustworthiness. We find that beliefs about others' cheating notions drive receiver behavior. We also find that, controlling for expectations about others' cheating notions, own cheating notions significantly affect cheating. We go further and provide novel evidence that parentally instilled values are significant determinants of cheating notions and that, moreover, different families of values that we label "cooperative" and "competitive" have opposite effects on individuals' notions of what constitutes cheating.

Finally, we provide a simple analytical framework explaining our findings. We show that the cheating behavior in our data can be rationalized by a model incorporating an intrinsic moral cost of cheating function in which both the decision whether to cheat or not as well

as the degree of cheating have negative utility consequences. In the process we provide evidence for, to our knowledge, a new explanation for the commonly-observed positive relationship between send amounts and trustworthiness: moral proscriptions against cheating the vulnerable.

An interesting question which we cannot address with our current data is how *knowing* that there are multiple notions of cheating affects sender and receiver behavior, either in the one-shot context here or when, more realistically, individuals interact repeatedly. One may wonder whether individuals adapt their own cheating notions to be more in line with the average population cheating notions causing an eventual convergence to one normative cheating standard; or, rather, whether those with high cheating notions cease to interact with the general population because they feel cheated more often in their interactions. We leave these and related questions for future research.

References

[1] Akerlof, George A. and William T. Dickens (1982), "The Economic Consequences of Cognitive Dissonance," *The American Economic Review*, 72, 307-319.

[2] Berg, J., Dickhaut, J. and K. McCabe (1995), "Trust, Reciprocity and Social History," *Games and Economic Behavior*, 10, 122-142.

[3] Bohnet, Iris and Richard Zeckhauser (2004), "Trust, Risk and Betrayal." *Journal of Economic Behavior and Organization*, 55(4), pp. 467-484.

[4] Bisin, Alberto and Thierry Verdier (2010), "The Economics of Cultural Transmission and Socialization." In Jess Benhabib, Alberto Bisin and Matthew O. Jackson editors: Handbook of Social Economics, Vol. 1A, The Netherlands: North-Holland, 2011, pp. 339-416.

[5] Castillo, Marco, Ragan Petrie, Torero Ragan, Maximo A. Torero and Lise Vesterlund (2012), "Gender Differences in Bargaining Outcomes: A Field Experiment on Discrimination." NBER Working Paper No. w18093

[6] Charness, Gary and Martin Dufwenberg (2006). "Promises and Partnership." *Econometrica*, 74(6), pp. 1579-1601.

[7] Charness, Gary and Matthew Rabin (2002). "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117(3), pp. 817-869.

[8] Chater, Nick, Steffen Huck and Roman Inderst (2010), "Consumer Decision-Making in Retail Investment Services: A Behavioral Economics Perspective", Report to the European Commission/SANCO.

[9] Cox, James C. (2004), "How to Identify Trust and Reciprocity," *Games and Economic Behavior*, 46, 260-281

[10] Dufwenberg, Martin and Gneezy, Uri (2000). "Measuring Beliefs in an Experimental Lost Wallet Game," *Games and Economic Behavior*, 30, pp. 163-182.

[11] Dufwenberg, Martin and Battigalli, Pierpaolo (2007). "Guilt in Games," *American Economic Review*, 97, pp. 170-176.

[12] Eagly, A.H. and M. Crowley (1986). "Gender and Helping Behavior: A Meta-Analytic Review of the Social Psychological Literature," *Psychological Bulletin*, 100, pp. 283-308.

[13] Eckel, Catherine C. and Philip J. Grossman (1998). "Are Women Less Selfish Than Men?: Evidence from Dictator Experiments," *Economic Journal*, 108, pp. 726-35.

[14] Ermisch, John and Diego Gambetta (2006). "People's trust: the design of a survey-based experiment," *ISER Working Paper Series 2006-34*, Institute for Social and Economic Research.

[15] Fehr, Ernst (2009), "On the Economics and Biology of Trust", *Journal of the European Economic Association*, 7, pp. 235-266.

[16] Fehr, Ernst and K.M. Schmidt (1999). "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*, 114 (3), pp. 817-868.

[17] Geanakoplos, John, David Pearce and Ennio Stacchetti (1989), "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, 1, pp. 60-79.

[18] Glaeser, Edward, David Laibson, Jose A. Scheinkman and Christine L. Soutter (2000), "Measuring Trust," *Quarterly Journal of Economics* 115(3), 811-846.

[19] Gneezy, Uri (2005), "Deception: The role of consequences," *American Economic Review*, March 2005, 384-394.

[20] Hung, Angela A., Clancy Noreen, Jeff Dominitiz, Eric Talley, Calude Berrebi and Farrukh Suvankulov (2008), "Investor and Industry Perspectives on Investment Advisers and Broker-Dealers", Technical Report, Rand Institute for Civil Justice.

[21] Inderst, Roman and Marco Ottaviani (2012), "Financial Advice," *Journal of Economic Literature*, 50(2): 494-512.

[22] Rabin, Matthew (1993), "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5), pp. 1281-1302.

[23] Reiss, Michelle C., and Kaushik Mitra (1998). "The Effects of Individual Difference Factors on the Acceptability of Ethical and Unethical Workplace Behaviors, *Journal of Business Ethics*, 17(14), pp. 1581-93.

[24] Ross, Lee, Greene, D., and House, P. (1977), "The False Consensus Phenomenon: An Attributional Bias in Self-Perception and Social Perception Processes," *Journal of Experimental Social Psychology*, 13(3), 279-301.

[25] Rousseau, Denise and Sim B. Sitkin and Ronald S. Burt and Colin Camerer (1998), "Introduction to Special Topic Forum: Not So Different After All: A Cross-Discipline View of Trust," *The Academy of Management Review*, 23(3), pp. 393-404.

[26] Sapienza, Paola, Anna Toldra and Luigi Zingales (2007), "Understanding Trust," NBER WP 13387

Table 1: Experimental design

	Number of sessions	Explicit cheating notion question opt-out	Investment fee	Max belief pay	Obs
Initial study	4	No	0.50 euro	5 euro	122
Additional sessions	4	Yes	0.50 euro (2 sessions) 0.00 euro (2 sessions)	20 euro	306

Table 2: Descriptive statistics

	Mean	Std Dev	Min	Max	N
Male	0.46	0.499	0	1	420
Age	23.73	4.171	18	58	420
Math score	7.66	1.251	3	10	402
Inc<30K	0.29	0.455	0	1	391
30≤Inc<45	0.24	0.426	0	1	391
45≤Inc<70	0.25	0.431	0	1	391
70≤Inc<120	0.16	0.366	0	1	391
Inc≥120K	0.07	0.249	0	1	391
Risk aversion	5.71	2.193	1	10	417
Send decision (binary)	0.81	0.392	0	1	428
Send amount	4.31	3.232	0	10	428
Average return proportion	1.28	0.697	0	4.02	427
Average expected return proportion	1.27	0.637	0	4.02	425
Competitive values emphasis	0.62	0.196	0	1	410
Good values emphasis	0.76	0.149	0.17	1	404
Expected probability of not being cheated	0.42	0.232	0	1	427
Average proportion of non-cheaters	0.49	0.376	0	1	428
Own cheating notion					

Table 3: Proportion of participants in sessions who opt-out of reporting a cheating notion, restricted to sessions with explicit opt-out opportunities "this has nothing to do with being cheated," by send amount

	Send Amount										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	Obs
	Proportion who selected "this has nothing to do with cheating"										
Mean	0.20	0.18	0.17	0.15	0.13	0.13	0.13	0.13	0.14	0.13	306
Std. Error	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
	Proportion who did not report a cheating notion for any reason										
Mean	0.23	0.21	0.21	0.17	0.15	0.15	0.15	0.16	0.17	0.17	306
Std. Error	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	

Notes: [1] In sessions with an explicit "opt-out" possibility participants could refrain from specifying an explicit personal cheating notion and instead respond either "I don't know" or "this has nothing to do with cheating." [2] The top row of Table 3 presents the proportion of participants who chose "this has nothing to do with cheating," while the lower row presents the proportion of participants who chose either of these two "opt-outs" or left the question entirely blank.

Table 4: Proportion of participants who feel cheated when receiving back less than sent

	Send Amount									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Sessions with "opt-out" possibility, restricted to those who could feel cheated									
Feel cheated if r<s	0.90	0.84	0.89	0.89	0.87	0.82	0.82	0.80	0.81	0.81
Std Err	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Obs	237	243	243	253	259	260	261	256	255	255
	Session without "opt-out" possibility									
Feel cheated if r<s	0.93	0.89	0.92	0.90	0.87	0.84	0.79	0.77	0.78	0.76
StdErr	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Observations	122	122	122	122	122	122	122	122	122	122
	All sessions pooled, restricted to those who reported a cheating notion									
Feel cheated if r<s	0.91	0.86	0.90	0.89	0.87	0.83	0.81	0.79	0.80	0.79
StdErr	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	359	365	365	375	381	382	383	378	377	377

Notes: [1] "Feel cheated if r<s" is a dummy that takes the value of one if a participant's threshold for feeling cheated conditional on sending *s* was greater than *s*. That is to say, this variable is an indicator for whether an individual would feel cheated if the return amount were strictly less than the send amount. [2] Robust standard errors, clustered by session, in parentheses. [3] The term "opt-out" refers to whether the question eliciting cheating notions had an explicit option to refuse to answer. In sessions with an opt-out, participants could choose either "I don't know" or "this has nothing to do with cheating" instead of specifying a personal cheating notion.

Table 5: Proportion of "equal-splitters."

	Send Amount									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Sessions with "opt-out" possibility, restricted to those who could feel cheated									
Proportion	0.35	0.32	0.29	0.22	0.29	0.32	0.33	0.30	0.30	0.35
Std Err	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Obs	237	243	243	253	259	260	261	256	255	255
	Sessions without "opt-out" possibility									
Proportion	0.30	0.33	0.34	0.27	0.30	0.34	0.33	0.26	0.29	0.33
StdErr	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Observations	122	122	122	122	122	122	122	122	122	122
	All sessions pooled, restricted to those who reported a cheating notion									
Proportion	0.33	0.32	0.30	0.24	0.29	0.32	0.33	0.29	0.30	0.34
StdErr	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	359	365	365	375	381	382	383	378	377	377

Notes: [1] "Equal-splitters" are participants who report they would feel cheated if they do not receive back at least half of the entire amount allocated to the receiver. Because experimental participants have a well-known predilection to state whole-number values, we label anyone whose definition of being cheated falls within the nearest whole-euro value of a precisely-equal split. For example, if a sender sends €1, a receiver receives €8.05, so we define equal-splitters for €1 sent to be anyone whose definition of being cheated falls within the interval [4, 5].

Table 6: Senders' decisions, Heckman estimates

	Main equation (1)	Selection equation (2)
Expected probability of not being cheated	2.76**	0.57
	(1.38)	(0.65)
Expected return from trusting	1.34***	0.28**
	(0.45)	(0.12)
(Probability of not being cheated) x(Expected return from trusting)	-1.57*	-0.07
	(0.85)	(0.46)
Low fee (dummy)	--	0.68***
		(0.09)
Age	0.11***	0.00
	(0.03)	(0.02)
Male	0.36	0.35**
	(0.32)	(0.14)
Math score	-0.00	0.12***
	(0.09)	(0.04)
Risk aversion	-0.14***	0.04
	(0.05)	(0.03)
Altruism	0.03	0.04
	(0.12)	(0.04)
30≤ Income <45	-0.29	0.13
	(0.42)	(0.25)
45≤ Income <70	-0.22	-0.04
	(0.59)	(0.23)
70≤ Income <120	-0.62**	-0.08
	(0.29)	(0.13)
Income ≥120	-0.63	0.74*
	(0.70)	(0.40)
Constant	1.45	-1.62***
	(2.16)	(0.61)
Obs	350	350
Mills Ratio	-.86	
	(0.36)	

Notes: [1] Robust standard errors, clustered by session, appear in parentheses. [2] *** = significant at 1%, ** = significant at 5%, * = significant at 10%. [3] For the Heckman model (cols 1-2): the dependent variable in the selection equation takes the value of 1 if the sender sends a positive amount and 0 otherwise; the dependent variable in the main equation is *how much* the sender sends. [4] The exclusion restriction for the selection equation consists of a dummy for "Low fee" sessions, a dummy taking the value of one if the observation came from a session where senders were charged nothing to send a positive amount, and 0 if the observation came from a session where senders were charged € 0.50 to send a positive amount [5] "Expected probability of not being cheated" is our measure of participants' subjective beliefs about not being cheated, described in the text. [6] "Expected return from trusting" is the participant's estimate of the proportion of money *sent* that receivers will return, averaged over all 10 possible send amounts. [7] "Risk aversion" is an index increasing in risk aversion obtained from an incentive compatible elicitation mechanism in a separate, unrelated, experiment. This variable takes values from 1 (risk loving) to 10 (very risk averse). [8] Altruism is how much emphasis participants' parents placed on the value "help others" during their upbringing. [9] Income variables refer to (self-reported) annual family income from all sources, in thousands of euros, net of taxes. The lowest category is excluded: "below 30 thousand euros".

Table 7: Intentional cheating by send amount

	Send Amount									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Own cheating notion	-0.08**	-0.13***	-0.08*	-0.07***	-0.04	-0.04*	-0.04	-0.05**	-0.03*	-0.04*
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Est. others' cheating notion	0.22***	0.21***	0.23***	0.22***	0.17***	0.15***	0.16***	0.17***	0.12***	0.14***
	(0.04)	(0.06)	(0.05)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Male	0.07	-0.02	0.18	0.06	0.03	-0.05	-0.16	-0.07	0.01	-0.06
	(0.07)	(0.16)	(0.12)	(0.14)	(0.20)	(0.15)	(0.14)	(0.13)	(0.13)	(0.18)
Age	-0.03	-0.04***	-0.03***	-0.02**	-0.02**	-0.03**	-0.01	0.01	-0.01	-0.01
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Math score	-0.04	-0.03	0.04	0.05	-0.04	-0.03	-0.06***	-0.06	-0.08**	-0.03
	(0.06)	(0.05)	(0.04)	(0.05)	(0.05)	(0.09)	(0.02)	(0.06)	(0.04)	(0.04)
Risk aversion	-0.00	0.03	0.01	-0.00	-0.02	-0.02	0.00	-0.01	-0.02	-0.07***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.02)
30≤ Inc	-0.02	0.18	0.24**	0.22	0.09	0.25*	0.50*	0.06	0.10	0.16
	(0.19)	(0.16)	(0.12)	(0.21)	(0.23)	(0.15)	(0.26)	(0.19)	(0.12)	(0.21)
45≤	0.12	0.01	0.06	0.10	0.29*	0.23***	0.43	0.24**	0.12	0.15
	(0.16)	(0.08)	(0.13)	(0.17)	(0.17)	(0.07)	(0.28)	(0.12)	(0.14)	(0.20)
70≤ Inc	0.17	0.17	0.07	-0.05	0.33*	0.41*	0.58**	0.70***	0.04	0.16
	(0.33)	(0.18)	(0.21)	(0.18)	(0.19)	(0.22)	(0.24)	(0.16)	(0.20)	(0.33)
Inc ≥120	0.00	-0.21	-0.07	0.01	-0.51	0.02	-0.21	-0.44	-0.04	-0.56*
	(0.35)	(0.28)	(0.21)	(0.20)	(0.32)	(0.28)	(0.40)	(0.31)	(0.29)	(0.33)
Constant	0.45	0.85	-0.69	-1.02*	-0.07	-0.10	-0.76*	-0.97	-0.05	-0.42
	(0.76)	(0.52)	(0.52)	(0.60)	(0.41)	(0.79)	(0.41)	(0.81)	(0.70)	(0.63)
Obs	369	366	366	369	371	370	371	369	366	366

Notes: [1] Each column presents estimates from a Probit model, with the (binary) dependent variable being "receiver intentionally cheats if sent relevant amount." Intentional cheating is defined by sending back strictly less than the receiver estimated senders needed back in order to not feel cheated. This threshold amount is also inserted as a control in each estimate by the variable "Est. others' cheating notion." [3] Robust standard errors, clustered by session, in parentheses. *** = significant at 1%, ** = significant ay 5%, * = significant at 10%. [4] Math score is individual's self-reported score on required math exams taken during the final year of high school in Italy. [5] Income variables refer to self-reported annual family income from all sources, in thousands of euros, net of taxes. The excluded category is "below 30 thousand euros annually". [6] Observations vary over columns because not all participants reported a cheating notion for every send amount. This is discussed in the text. Additionally, we do not have demographics for all participants.

Table 8: Determinants of cheating notions

Dependent variable = Own cheating notion										
Competitive values	Competitive values	€ sent	(€ sent)^2	Male	Age	Math score	Risk aversion	Cons	Obs	Individuals
-2.55**	1.63**	1.07***	-0.02***	-0.47	0.00	-0.02	-0.11	3.55***	3496	354
(1.09)	(0.64)	(0.07)	(0.01)	(0.43)	(0.03)	(0.11)	(0.08)	(1.31)		

Notes: [1] Estimates are from an individual-level random effects regression model. [2] Variables present in the regression, but omitted for readability: full set of income dummies; dummy for sessions with no investment fee; dummy for sessions comprising the initial study. None of these variables had significant coefficients. [3] Robust standard errors, clustered by session, appear in parentheses.

Table 9: Sensitivity of amounts returned to cheating notions by decision to cheat, Heckman models

	Send Amount									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Conditional on not cheating (euros returned >= estimated others' cheating notion)										
Est. others' cheating notion	1.17***	1.02***	0.97***	1.19***	1.07***	0.95***	0.88***	0.89***	1.09***	1.02***
	(0.17)	(0.13)	(0.14)	(0.25)	(0.15)	(0.11)	(0.16)	(0.13)	(0.22)	(0.13)
Constant	3.83**	4.98**	3.54**	4.68*	5.06**	4.88***	5.96***	4.97**	8.15**	9.26***
	(1.95)	(2.25)	(1.73)	(2.78)	(2.39)	(1.85)	(2.10)	(2.00)	(4.06)	(3.02)
Demographics	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	311	319	320	328	333	334	335	332	329	329
Wald test: *est. others' cheating notion* coefficient = 1 (p-value)										
	0.32	0.86	0.84	0.43	0.63	0.66	0.43	0.39	0.67	0.84
Conditional on cheating (euros returned < estimated others' cheating notion)										
Est. others' cheating notion	0.42***	0.37***	0.57***	0.38***	0.44***	0.43***	0.49***	0.58***	0.63***	0.53***
	(0.07)	(0.06)	(0.10)	(0.10)	(0.10)	(0.09)	(0.13)	(0.12)	(0.14)	(0.12)
Constant	-0.16	0.93	0.18	0.95	1.68	0.03	1.09	0.09	-0.38	-0.01
	(0.92)	(1.02)	(1.51)	(1.92)	(1.91)	(2.01)	(2.87)	(3.02)	(3.36)	(3.31)
Demographics	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	311	319	320	328	333	334	335	332	329	329
Wald test: *est. others' cheating notion* coefficient = 0.5 (p-value)										
	0.23	0.04	0.47	0.24	0.52	0.43	0.95	0.52	0.34	0.79

Notes: [1] Standard errors in parentheses. *** = significant at 1%, ** = significant at 5%, * = significant at 10%. [2] Each column presents a Heckman model estimate using as its exclusion restriction participants' own cheating notions. [3] The dependent variable in column i is the amount a participant will send back if the sender sends i euros, i=1,…,10. [4] The reported independent variables in column *i* are: "Est others' cheating notion" is each participant's estimate of the minimum amount of money a sender would need back in order to not feel cheated when the sender sends *i* euros, i=1,…,10. [5] Each estimate includes our standard set of demographic controls, omitted for readability from the table. These controls are: gender, age, math score, family income and risk aversion.

Figure 1: Bimodal cheating definition distributions, for send amount 1 euro

Figure 2A: Bimodal cheating definition distributions (sessions with no "opt-out")

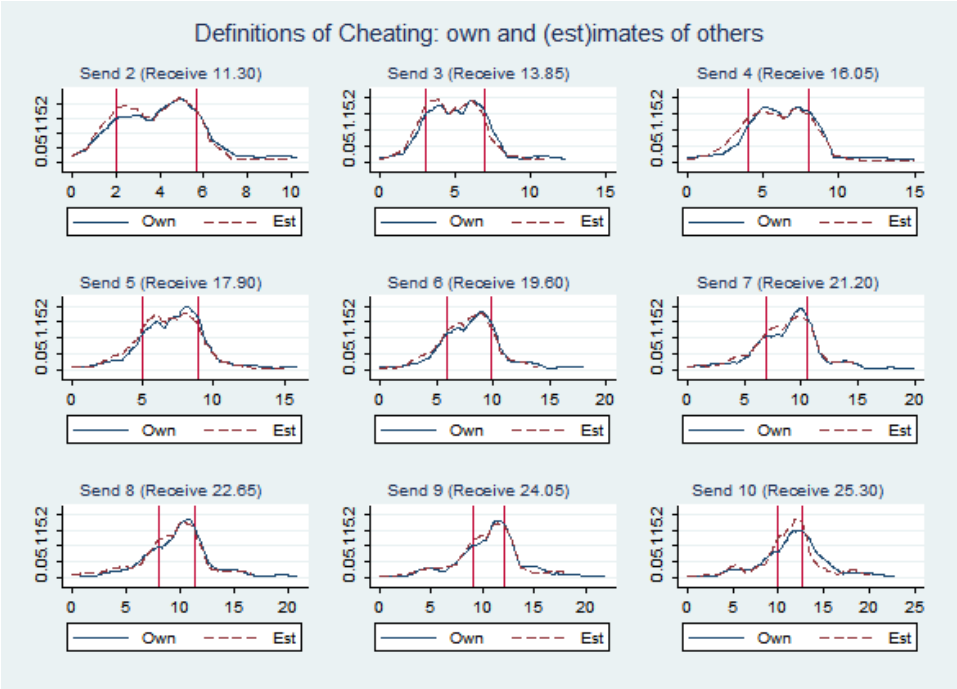


Figure 2B: Bimodal cheating definition distributions (sessions with "opt-out.")

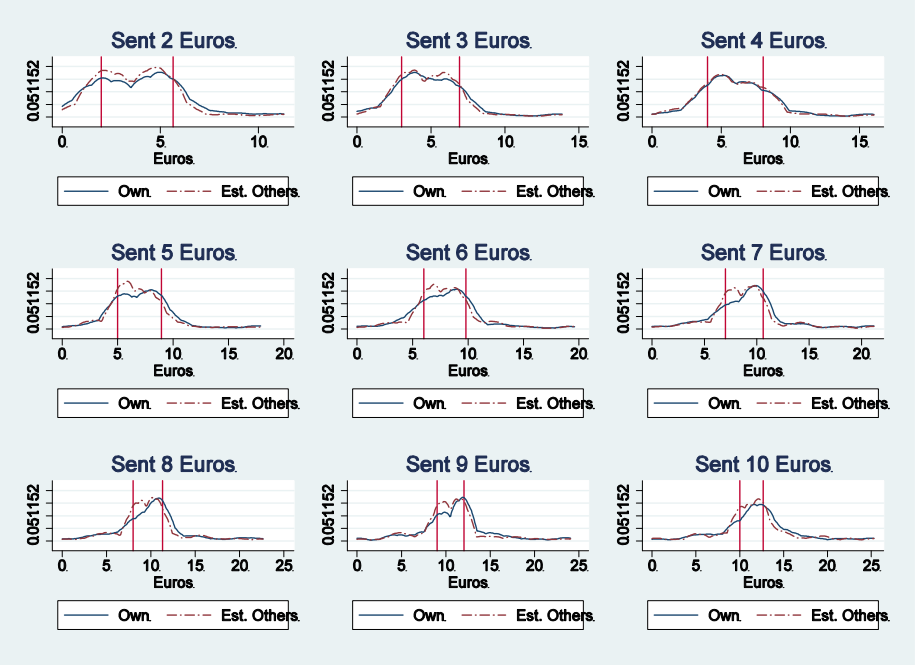


Figure 2C: Bimodal cheating definition distributions, all sessions pooled

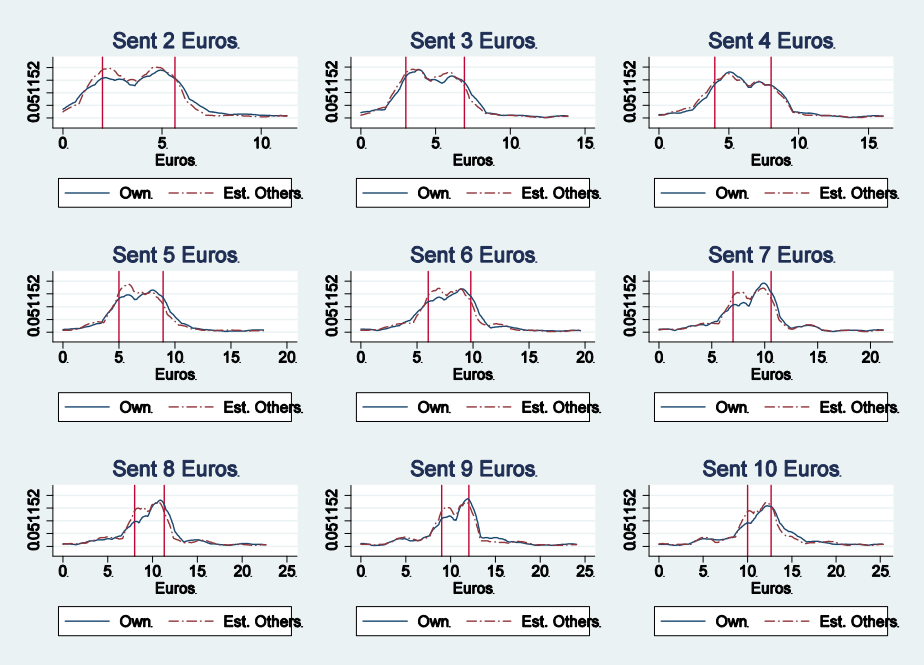
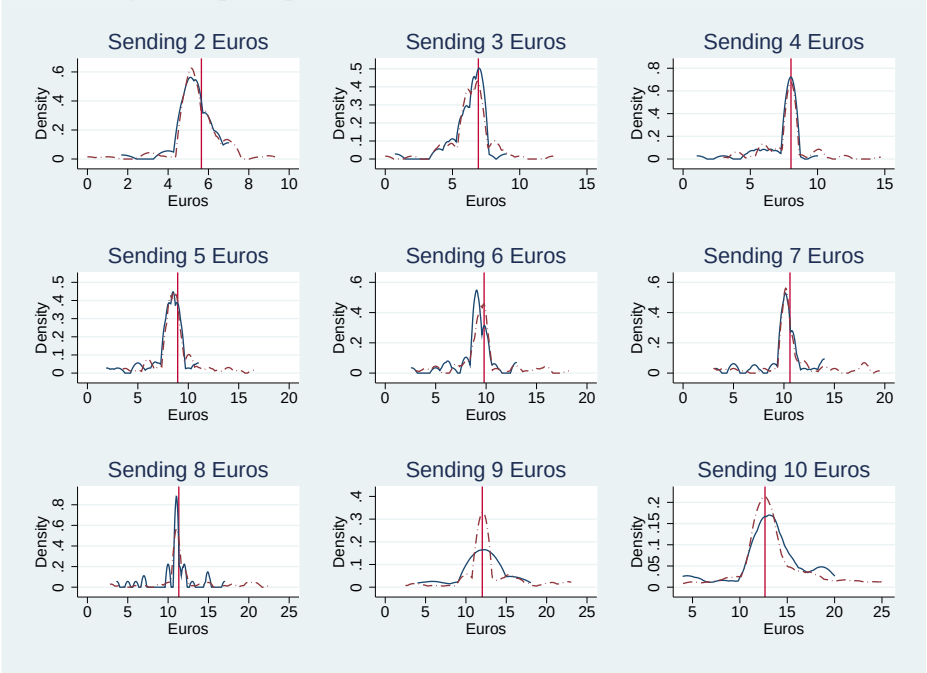
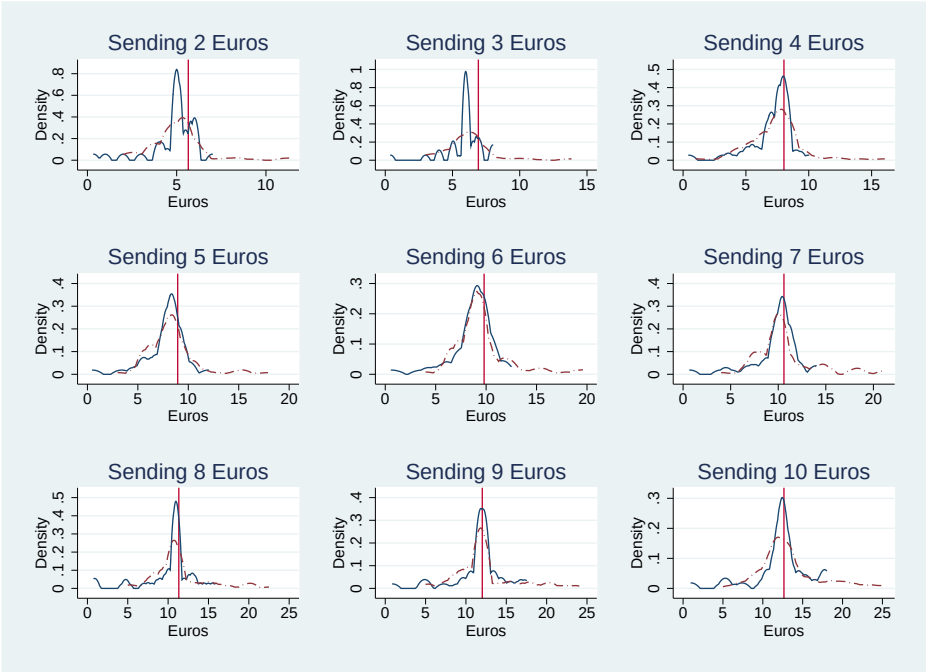


Figure 3A: Consistency of "equal splitters" defined for send amount 1, across other send amounts.



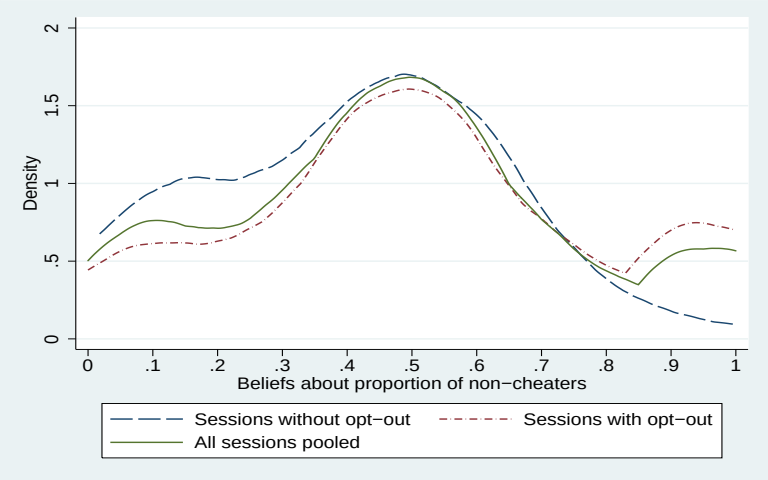
Notes: [1] The figure presents kernel density plots of cheating notions for send amounts 2-10 for those labeled as equal splitters when considering sending 1 euro. [2] Dashed lines represent data from the additional sessions, where opt-out was possible, while solid lines represent data from the initial study. [3] The vertical bars represent half of the total amount receivers if senders send the associated amount.

Figure 3B: Consistency of "equal splitters" beliefs



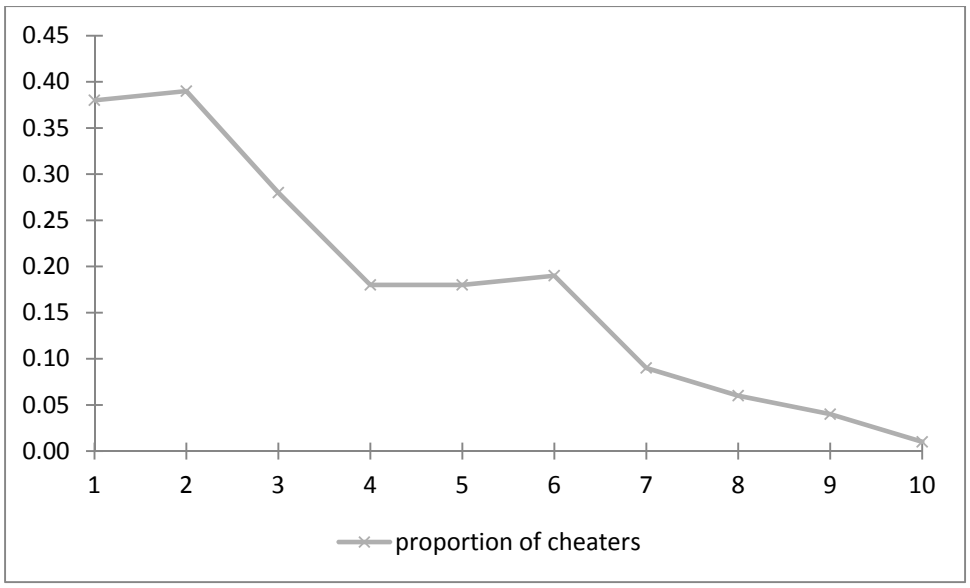
Notes: [1] The figure presents kernel density plots of beliefs about others' cheating notions for send amounts 2-10 for those whose own cheating notions classify them as equal splitters when considering sending 1 euro. [2] Dashed lines represent data from the additional sessions, where opt-out was possible, while solid lines represent data from the initial study. [3] The vertical bars represent half of the total amount receivers if senders send the associated amount.

Figure 4, Beliefs about the probability of not being cheated



Notes: Observations in the sessions with opt-out (short-dash line) are restricted to individuals who have a cheating notion for every possible amount a sender could send. This is to ensure our summary measure of beliefs about the probability of being cheated is well-defined. Thus the density plot for the additional sessions is based on 207 (out of 306) observations.

Figure 5: Proportion of cheaters by send amount



Notes: The figure reports the proportion of cheaters (y-axis), after partialling out the effect of expectations of others' cheating notions, for each possible send amount (x-axis).

Not for publication

Appendix I: Robustness Checks

A Additional Robustness check treatments

In addition to our main experiment described in Appendix II, two further treatments were conducted for robustness. First of all, to check whether there is something peculiar about the on-line environment driving our results or whether paying only 10 percent of participants provides incentives that are too weak, we ran two sessions in the laboratory where 100 percent of participants were paid. As a second robustness exercise, we conducted sessions in which our direct cheating notion question was omitted and replaced with a series of questions asking participants how they would feel about various possible outcomes in the trust game from the point of view of the sender. The purpose of this latter treatment is to address the concern that our direct cheating notion question might prime participants to associate cheating with the trust game.

A.1 In-lab sessions

In total, 36 individuals took part in two sessions conducted in the experimental laboratory at the Einaudi Institute for Economics and Finance in Rome, Italy. Participants were recruited from the same subject pool as were the on-line sessions. There was no overlap in actual participants—i.e., no participant took part in both an on-line session and an in-lab session. All in-lab participants were paid based on their choices in the experiment and the accuracy of the their reported beliefs.

Apart from taking place in the laboratory, the design of this treatment and the materials used were exactly the same as the on-line treatments. Participants simply completed the on-line experiment in the laboratory. All sessions of the in-lab experiment allowed participants to opt out of specifying a cheating notion by selecting one of two responses: "I don't know" or "this has nothing to do with cheating." Neither session featured a fee to send a positive amount.

Participants' return proportions, cheating notions and beliefs were remarkably similar across the on-line and in-lab environment (Table A1). On the other hand, in-lab senders were slightly more likely to send a positive amount than their on-line counterparts, raising the average amount sent by in-lab senders. Conditional on sending a positive amount, average send amounts were similar: 5.36 in on-line low fee sessions; 5.43 in the laboratory; with standard errors 0.25 and 0.44, respectively.

Next, consider how send amounts vary with cheating and monetary return beliefs (Table A2). Because we have few observations and lack the exogenous variation in senders' incentives which we exploited in the analysis of our on-line data, we account for selection into sending a positive amount here by estimating a Tobit model rather than a Heckman model. The results paint a picture qualitatively similar to the on-line data: amounts sent vary positively and significantly with both expected (lack of) cheating and expected return.

In terms of cheating notions, the picture is also quite similar in the lab and on-line experiments: the vast majority of participants have a cheating notion for all possible send

amounts (Table A3); the vast majority have a cheating notion demanding at least a positive return on investment (Table A4); roughly one-third of participants are "equal-splitters" (Table A5). Furthermore, as in the on-line data, own cheating notions are negatively related to intentional cheating while expectations about others' cheating notions are positively related to intentional cheating (Table A6).

A.2 Treatments without cheating notion question

We also conducted (on-line) sessions of a treatment in which we dropped our direct cheating notion question and replaced it with a section where participants were asked to indicate how they would feel, as a sender, about various send/return amount scenarios. In total, 170 participants took part in this treatment. As with the main study, ten percent of participants were randomly chosen to be paid their experimental earnings.

To keep the number of individual questions reasonable, we selected three common send amounts—$S = 1, 5$ and 10—and, for each of these, asked participants how they would "feel" if the receiver returned four specific amounts: $0, \frac{S}{2}$, S and $\frac{f(S)}{2}$. These send/return scenarios were chosen to line up with the cheating notions common in the data from our main study. In terms of feelings, for each send/return amount scenario participants were asked to select exactly two options from a list of several options that best described how they would feel if the scenario were realized. The list of options included positive evaluations ("[the receiver] was generous," "[the receiver] treated me fairly"), neutral evaluations ("[the receiver] was intelligent," "I have no particular opinion of [the receiver's] behavior") and negative evaluations ("[the receiver] cheated me," "[the receiver] disappointed me"). A free-form response option was also available.

To compare the qualitative data we have in this treatment with data from our main sessions, for each send/return scenario investigated in this treatment we calculate the proportion of participants in our main treatment who would feel cheated according to their own reported cheating notions. We compare this proportion to the proportion of respondents in the "feelings" treatment reporting feeling "disappointed" or "cheated." To maximize comparability, from our main treatment data we use only sessions where participants were allowed to opt out of specifying a cheating notion. We find a strong positive relationship between the proportion of participants expressing negative feelings in particular scenarios and the implied proportion of participants feeling cheated in those scenarios in the data from the main treatment (Figure A1). We interpret this as support for the view that trust game participants have well-defined cheating notions and evidence against the view that the cheating notions they report can be mainly attributed to priming.

A.2.1 Evidence on receivers' motivations

In sessions without a direct cheating notion question, at the end of the experiment we added a section in which participants were asked to descibe the rationale they used, if any, for deciding how much to return in the role of receiver. Participants were asked:

> *Describe, in general, how you arrived at your decisions concerning how much to return when you played role B [receiver] for each amount A could have sent you*

Participants could select among four pre-programmed options, or, if none on the list suited them they could select "other" and specify their own rationale. Three of the four pre-programmed responses were meant to capture positive reciprocity, ("the more A [the sender] sent, the more I returned in order to reward nice behavior"); negative reciprocity ("the less A [the sender] sent, the less I returned, in order to punish bad behavior"); vulnerability ("the more A [the sender] sent, the more I returned in order to compensate A [the sender] for being at the mercy of my actions"). The fourth pre-programmed option was essentially a decline to state option ("I did not have any particular rationale in mind.").

Table A7 presents the results. Overall, 83 percent of participants selected one of the four pre-programmed option. The modal response, selected by 42 percent of participants, was that receivers return more when senders send more to compensate senders for their vulnerability. The second most common response reflected positive reciprocity. Almost nobody (6 percent) selected negative reciprocity as their primary rationale, while a similarly low percentage selected the pre-programmed decline to state option (6 percent).

B Robustness checks on beliefs

A common concern whenever beliefs are elicited is the extent to which the elicitation mechanism itself colors reported beliefs. Monetary incentives meant to ensure that participants report beliefs truthfully may give rise to other potential confounds, such as hedging motives: by shading reported beliefs toward bad outcomes, individuals may reduce the variance of their experimental earnings. On the other hand, monetary incentives that are too weak can allow reported beliefs to be non-truthful for various reasons. In particular, one may worry that the significant correlation between estimates of others' cheating notions and own return amounts arises because of a tendency for participants to ex-post rationalize their receiver strategies: by reporting believing that whatever they return is enough to not cheat others, participants can maintain a positive moral self-image.

First we consider ex-post rationalization. If ex-post rationalization is driving beliefs about others' cheating notions, then quadrupling the incentives for belief accuracy in the additional sessions should make this motive less relevant. Therefore, evidence of ex-post rationalization in initial sessions would be a consistently smaller correlation between return amounts and beliefs about others' cheating notions in the "high belief pay" sessions. Table A8 (panel A) presents panel regressions of beliefs about others' cheating notions as a function of return amounts incorporating a dummy for high belief pay and an interaction with return amounts. The coefficient of interest is on the interaction between high belief pay and return amount: if ex-post rationalization is important when belief pay is low, and diminished for high belief pay, we would expect this coefficient to be consistently negative and significant. Instead, the estimated coefficient on the interaction term is positive and (marginally) significant providing evidence against ex-post rationalization. Adding our standard set of demographics does not change the results. Moreover, restricting to the subset of observation where the receiver does not cheat—where the ex-post rationalization argument has the most bite—changes nothing qualitatively.[1]

[1] We omit these last two robustness checks to save space, but they are available on request. It should also be noted that variation in belief pay could not have direcly affected receivers' actions, since participants did

Next, consider hedging motives. As a concrete example, consider a sender who has chosen to send 10 euros. If the sender believes the receiver is trustworthy and reports this belief, then in the good state of the world where the receiver *is* trustworthy, the sender earns a lot—both beliefs and actions pay off. However, in the bad state of the world where the receiver returns nothing, the sender loses quite a lot—neither actions nor beliefs pay off. By shading reported beliefs downward—towards a higher likelihood of an untrustworthy sender—the sender can shift some earnings out of the good state of the world into the bad state of the world, reducing earnings variance, i.e., risk.

To test for hedging motives in beliefs, we estimate participants' stated beliefs about the amount of money receivers will return for each possible send amount. We present panel regressions, where we control for whether a sender actually chose to send a particular amount, risk aversion, an interaction between these two variables. Since hedging motives can only (literally) apply to the send amount a sender actually chose to send, one measure of the hedging motive is the coefficient on the dummy for actually chosen send amounts. A secondary prediction is that more risk averse individuals care about hedging more, so the interaction term should be negative. Table A8 (panel B) presents our estimates, which provide no support for the importance of hedging. In fact, contrary to hedging motives, reported beliefs about return amounts are (marginally) significantly *higher* for the amount a sender actually chose to send as evidenced by the coefficient on "Chosen send amount." Risk aversion plays no significant role. Controlling for demographics and/or the level of belief pay does not change anything qualitatively.

not know there would be a belief elicitation section until after they had submitted their strategies.

Table A1: Comparison of behavior in the lab and on-line, summary statistics

	Send > 0	Send amount	Return proportion	Return proportion (belief)	Proportion of non-cheaters	Proportion of non-cheaters (belief)
			In-lab sessions			
	0.97	5.28	1.25	1.36	0.43	0.56
	(0.03)	(0.45)	(0.10)	(0.10)	(0.06)	(0.03)
Obs	36	36	36	36	36	36
			On-line low fee sessions			
	0.90	4.83	1.28	1.22	0.53	0.53
	(0.03)	(0.26)	(0.06)	(0.06)	(0.03)	(0.02)
Obs	150	150	149	148	150	135

Table A2: Send amount (Tobit), in-lab sessions

	Dependent variable = send amount		
	(1)	(2)	(3)
Expected probability of not being cheated	4.29*	4.94**	6.97***
	(2.19)	(2.25)	(1.90)
Expected return from trusting	1.54*	1.57**	1.41*
	(0.81)	(0.75)	(0.77)
Male		1.53*	0.93
		(0.81)	(0.75)
Age		-0.16*	-0.30**
		(0.09)	(0.11)
Math score		-0.34	0.07
		(0.42)	(0.37)
Risk aversion			-0.47**
			(0.17)
Altruism			0.04
			(0.21)
30≤ Income <45			-1.48
			(0.98)
45≤ Income <70			0.03
			(1.10)
45≤ Income <70			1.76
			(1.47)
Income ≥120			-2.99**
			(1.26)
Constant	0.86	5.85	8.66*
	(1.52)	(4.60)	(5.01)
Obs	36	34	32

Notes: [1] Robust standard errors in parentheses. [2] *** = significant at 1%, ** = significant at 5%, * = significant at 10%. [3] Each column presents a Tobit model estimate where the dependent variable is *how much* the sender sends and censoring below 0 is taken into account. [5] "Expected probability of not being cheated" is our measure of participants' subjective beliefs about not

being cheated, described in the text. [6] "Expected return from trusting" is the participant's estimate of the proportion of money *sent* that receivers will return, averaged over all 10 possible send amounts. [7] "Risk aversion" is an index increasing in risk aversion obtained from an incentive compatible elicitation mechanism in a separate, unrelated, experiment. This variable takes values from 1 (risk loving) to 10 (very risk averse). [8] Altruism is how much emphasis participants' parents placed on the value "help others" during their upbringing. [9] Income variables refer to (self-reported) annual family income from all sources, in thousands of euros, net of taxes. The lowest category is excluded: "below 30 thousand euros".

Table A3: Proportion of participants with a cheating notion, in-lab sessions

	Send amount									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Proportion w/ cheating notion	0.72	0.86	0.83	0.92	0.94	0.94	0.97	0.97	0.97	0.97
	(0.08)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Obs	36	36	36	36	36	36	36	36	36	36

Notes: [1] Raw proportions reported. [2] Standard errors appear in parentheses

Table A4: Proportion of participants who would feel cheated by (return amount)<(send amount), in-lab sessions

	Send Amount									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Proportion w/ (cheating notion) ≥ (send amt)	0.88	0.84	0.97	0.94	0.94	0.97	0.89	0.83	0.83	0.86
	(0.06)	(0.07)	(0.03)	(0.04)	(0.04)	(0.03)	(0.05)	(0.06)	(0.06)	(0.06)
Obs	26	31	30	33	34	34	35	35	35	35

Notes: [1] Reported proportions are conditional on specifying a cheating notion. [2] Standard errors appear in parentheses

Table A5: Proportion of participants with an "equal-split" cheating notion, in-lab sessions

	Send Amount									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Proportion of "equal splitters"	0.35	0.23	0.20	0.18	0.32	0.32	0.34	0.23	0.23	0.37
	(0.10)	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.08)
Obs	26	31	30	33	34	34	35	35	35	35

Notes: [1] Reported proportions are conditional on specifying a cheating notion. [2] "Equal splitters" are defined by having a cheating notion, c, such that floor(f(s)/2) ≤ c ≤ ceiling(f(s)/2), where floor(x) is the least integer less than x and ceiling(x) is the least integer greater than x. This takes into account the well-known predilection of experimental participants for integer values. For example, since floor(4.025) = 4 and ceiling(4.025) = 5, a participant is labeled an "equal splitter" for send amount 1 if his or her cheating notion, c, satisfies 4 ≤ c ≤ 5. [3] None of the reported proportions are significantly different from 0.30 at the 10 percent significance level using two-tailed difference in proportions tests.

Table A6: Intentional cheating (reduced form), in-lab sessions

	Sent Amount									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Own cheating notion	-0.78*	-0.35**	-0.08	-0.05	0.02	-0.08	-0.05	-0.02	-0.25**	-0.13*
	(0.43)	(0.17)	(0.12)	(0.11)	(0.07)	(0.09)	(0.11)	(0.11)	(0.12)	(0.07)
Estimate of others' cheating notions	1.08**	0.32	0.15	0.16	0.18	0.05	0.25**	0.11	0.28*	0.34***
	(0.50)	(0.20)	(0.17)	(0.14)	(0.13)	(0.13)	(0.11)	(0.13)	(0.16)	(0.11)
Constant	-0.68	0.27	-0.36	-0.73	-1.13	0.30	-1.49	-0.41	-0.11	-1.98*
	(0.64)	(0.62)	(0.74)	(0.87)	(0.94)	(0.85)	(1.09)	(1.06)	(1.01)	(1.11)
Obs	26	31	30	33	34	34	35	35	35	35

Notes: [1] Each column presents estimates from a Probit model, with the (binary) dependent variable being "receiver intentionally cheats if sent relevant amount." Intentional cheating is defined by sending back strictly less than the receiver estimated senders needed back in order to not feel cheated. This threshold amount is also inserted as a control in each estimate by the variable "Estimate of others' cheating notion." [3] Robust standard errors, clustered by session, in parentheses. *** = significant at 1%, ** = significant ay 5%, * = significant at 10%.

Table A7: Proportion of receivers specifying a particular rationale

	Overall	High fee sessions	Low fee sessions
Sender vulnerability	0.42	0.40	0.45
	(0.04)	(0.05)	(0.06)
Positive reciprocity	0.29	0.31	0.27
	(0.04)	(0.05)	(0.05)
Negative reciprocity	0.06	0.06	0.05
	(0.02)	(0.03)	(0.03)
No motive	0.06	0.05	0.08
	(0.02)	(0.02)	(0.03)
Obs	170	93	77

Notes: [1] Raw proportions reported; [2] Standard errors in parentheses; [3] Proportions in each column sum to less than one, with the unaccounted for observations being participants who elected to supply their own rationale rather than one of the four pre-programmed rationale; these self-supplied rationale varied widely and are not easily classifiable.

Table A8: Robustness checks, main study data

Panel A: checking for ex-post rationalization

Dependent variable = Expected others' cheating notion

Return amount	Amount sent	High belief pay	(High belief pay) X (Return amt)	Cons	Obs	Individuals	R^2
0.11***	0.85***	0.12	0.05*	1.58***	4254	428	0.5
(0.02)	(0.03)	(0.24)	(0.03)	(0.20)			

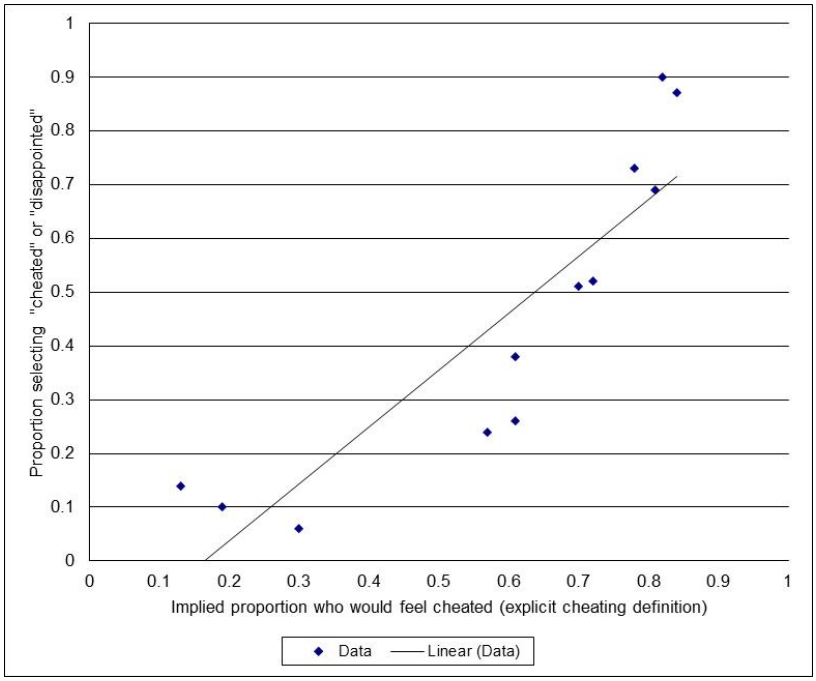
Panel B: checking for hedging motives in beliefs

Dependent variable = Expected return amount

Amount sent	Chosen send amount	Risk aversion	(Chosen send amt) X (Risk aversion)	Cons	Obs	Individuals	R^2
0.82***	0.29*	-0.00	-0.02	1.61***	4146	417	0.34
(0.02)	(0.17)	(0.04)	(0.03)	(0.23)			

Notes: [1] Both the top and bottom panel report individual random effects regressions pooling observations across all send amounts. [2] Robust standard errors, clustered by session, appear in parentheses. [3] "High belief pay" is a dummy taking the value of one if the session involved a 20 euro maximum belief pay, and 0 if the maximum possible belief pay was 5 euros; "Chosen send amount" is a dummy variable indicating the amount a participant actually chose to send in the role of sender; "Risk aversion" is an incentive-compatible index of risk aversion obtained from a previous experiment. [4] We drop observations for which we have no measure of risk aversion.

Figure A1: Comparison of proportion feeling cheated by elicitation method



Appendix II: Experiment Instructions

In this experiment, you will be randomly paired with another participant and assigned randomly one of two roles: A or B. This pairing will be anonymous. Neither the person in the role of A nor the person in the role of B wil know with whom they have been paired.

The role of A

The player in the role of A is given 10.50 euros and must decide whether to send some all or none of this money to the player in the role of B, the person with whom A has been paired. [If A decides to send some of the this money, A will be charged a fee of 0.50 euros.] For every euro that A sends, B will receive more than 1 euro according to the table below.

If A sends €	1	2	3	4	5	6	7	8	9	10
B receives €	8.05	11.3	13.85	16.05	17.9	19.6	21.2	22.65	24.05	25.3

The role of B

After A makes his or her decision about how much to send to B, B decides how much of the money he or she receives—the amounts in the table above (8.05 euros, 11.30 euros, etc.)—to return to A. The player in the role of B will specify an amount to return for each possible amount they could receive. For example, if A sends 4 euros and B therefore receives 16.05 euros, B must decide how much of this 16.05 euros to return to A; and a decision must be made for every amount A could send (1,2,3,. . . ,10 euros).

Your earnings

For every pair of participants, one in the role of A and one in the role of B, the decisions that both A and B make determine the pairs earnings. Both A and B will be informed of the outcome determined by their choice.

In general:

- If A sends a positive amount to B:
    1. A's earnings will be: € 10.50 – (euros sent to B) + (euros returned by B) – (€ 0.50 fee)
    2. B's earnings will be: (euros received by B according to the table above) – (euros returned to A)
- If A sends nothing to B:
    1. A's earnings will be € 10.50
    2. B's earnings will be € 0.

Specifically, for every pair of players the result of this situation will be determined as follows:

i Every participant specifies their decision for each possible role (A and B).

ii The computer will randomly assign a role to each participant and randomly and anonymously pair each participant assigned the role of A with a participant assigned the role of B.

iii Within each pair, A's decisions will be combined with B's decision to determine the outcome for both A and B.

A Experiment Screens

A.1 Sender decision screen 1

If you are assigned the role of A, do you want to send money to B? If you send money, you will be charged a € 0.50 fee.

Choose "send" or "don't send" on this screen. If you choose "send", you will specify the amount to send on the next screen.

__ Send money
__Don't send money

A.2 Sender decision screen 2

How much money do you want to send if you are assigned the role of A?

__ € 1
__ € 2
...
__ € 10

A.3 Receiver decision screens

[There are 10 separate screens. A representative question is below.]

Imagine that you have been assigned the role of B ...
How much will you send back to A if A sends € 7 and you therefore receive € 21.20?

A.4 Cheating definition screen

If you are assigned the role of A, what is the minimum amount you would need to receive back from B in order to not feel cheated?

If you send €1 and therefore B receives €8.05, you would need back : ____

Insert a number above, or select one of the two following options:
__ This has nothing to do with cheating

__ I do not know

...
If you send €10 and therefore B receives €25.30, you would need back : ____

Insert a number above, or select one of the two following options:
__ This has nothing to do with cheating
__ I do not know

A.5 Belief elicitation

A.5.1 Instructions, screen 1

Now, we begin a new section. In this section as in the previous section, each question can contribute to your potential earnings.

Specifically, in this section you will be asked to estimate the choices other participants made in the previous section. Every question is about the choices of other participants, so please exclude your own actions from your estimations. The accuracy of your estimates will be calculated excluding your own actions as well.

Your earnings from this section will be determined by choosing one of your estimations at random and paying you according to the accuracy of this randomly chosen estimation. Every estimate has the same chance of being chosen by the computer. Your potential earnings from this experiment will be the sum of your earnings in this section and in the previous section.

The formula used to calculate your earnings from the randomly-chosen estimate is detailed on the next page.

A.5.2 Belief compensation formula screen

The method used to calculate your earnings from your estimates is detailed below. The most important thing to notice is that more accurate estimates have higher chances of earning money.

- Your estimate, R, is inserted into the following formula where "r" stands for the true value of the thing being estimated and "r_{max}" is the maximum value this true value can attain.

$1 - \left(\frac{R-r}{r_{\max}}\right)$

- This produces a number between 0 and 1. Call this number "z".
- The computer chooses a number between 0 and 1 with each number in between 0 and 1 being equally likely. Call this number "y".

- If $y \leq z$, you will earn €5.00 [€20.00] for your estimate.
- If $y > z$, you will earn €0.00 for your estimate.

An example

Suppose you are asked to estimate the average amount participants in the role of A send in the previous section of this experiment. And, imagine that this average turns out to actually be €4.00. The maximum value this average could have taken is €10. Therefore "r_{max}" in the equation above is 10 and r is 4. The equation therefore becomes:
$1 - \left(\frac{R-4}{10}\right)$
Notice that the closer your estimate, R, is to the actual value of 4 in our hypothetical example, the larger is z and therefore the larger is the probability of earning €5 [€20.00] for your estimate rather than €0.

- If your estimate is exactly correct, then (R-4)/10 = 0 and therefore z=1. Because the number chosen by the computer is at most one, an exactly correct estimate always pays €5 [€20.00].
- On the other hand, the probability with which your estimate earns you €5 [€20.00] diminishes the farther away from the true value your estimate is: z becomes smaller and so does the chances that $y < z$.

Click continue to begin start the estimation section

A.5.3 Beliefs elicitation screen 1

How much, on average, will players in the role of A send to B's? Insert a number between 0.00 and 10.00 : _ _ _

A.5.4 Beliefs elicitation screen 2

How much, on average, will B's return to A's?

If A sends €1 and B therefore receives €8.05, B's will return on average: _ _ _
...
If A sends €10 and B therefore receives €25.30, B's will return on average: _ _ _

A.5.5 Beliefs elicitation screen 3

What is the minimum amount (on average) that A's will need back from B's in order to not feel cheated?

If A sends €1 and B therefore receives €8.05, to not feel cheated A will need back from B at least: _ _ _

...

If A sends €10 and B therefore receives €25.30, to not feel cheated A will need back from B at least: ___

A.5.6 Beliefs elicitation screen 4

What percent of participants in the role of B will return enough money to you (if you are assigned the role of A) so that you don't feel cheated?

If you send €1 and B therefore receives €8.05, what percent of B's will return enough so that you don't feel cheated?: ___

...

If you send €10 and B therefore receives €25.30, what percent of B's will return enough so that you don't feel cheated?: ___

A.5.7 Beliefs elicitation screen 5

How much money (on average) do other participants in the role of A believe will be returned to them by B's?

If A sends €1 and B therefore receives €8.05, how much money does A believe B will return? ____

...

If A sends €10 and B therefore receives €25.30, how much money does A believe B will return? ____