Cheating in the Trust Game*

Jeffrey V. Butler
EIEF

Paola Giuliano
UCLA, NBER and IZA

Luigi Guiso
European University Institute, EIEF and CEPR

June 13, 2010

Abstract

A large experimental literature has investigated the determinants of trust using variants of the so called trust game. However, a lot of ambiguity remains regarding the definition of what one is trusting: one extreme view is that since there are no explicit promises or guarantees made in the trust game, cheating is not defined and therefore trust is not implied. We address this concern by asking trust game participants their notion of cheating. We find that i) senders do have a notion of cheating, ii) the 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. Once the notion of cheating is defined, we go further and show that participants' decision to trust at all is significantly affected by their beliefs about the probability of being cheated. Additionally, we look at receivers' decisions to cheat and uncover a number of new results. For example, we show that values instilled by parents affect cheating behavior, and that the impact of values is magnified by counterparty vulnerability.

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

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

Very Preliminary.

Comments are welcome.

*We thank many people.

1 Introduction

A large experimental literature has investigated the determinants of trust using variations of the so-called trust game[1]. The trust game is a sequential-moves game of perfect information involving two players: a "sender" and a "receiver." The sender moves first by deciding whether to send some, all or none of a fixed endowment to the receiver. Any positive amount sent is multiplied by a certain amount by the experimenter. The receiver moves second and decides whether to return some, all or none of this (multiplied) amount to the sender. Sending a positive amount is often interpreted as an expression of *trust*, since receivers cannot commit to return any particular sum. Returning a positive amount is typically referred to as *trustworthiness*, since receivers' pecuniary incentives dictate returning nothing (see Berg, Dickhaut and McCabe, 1995).

Despite the popularity of the trust game[2], a lot of ambiguity still remains regarding the definition of what one is trusting. Is the sender trusting the receiver to return at least what he sent? Or is he trusting the receiver to send back a good chunk of the surplus that is created by the act of trusting?[3] A common concern is that because receivers neither make any promise nor enter into any explicit agreement specifying what they should do with their windfall, there is no scope for senders to feel cheated by receivers' actions and hence no role for trust in the trust game at all[4].

In this paper we contribute to the trust game debate by asking whether participants have a notion of "cheating" even when no explicit promise is made and whether this notion affects their behavior in the trust game. A positive answer reassures that trust in the trust game does indeed capture what one is trying to measure: the sender's "....intention to accept vulnerability based upon positive expectations of the intentions or behavior of others" using the definition justified in Rousseau et. al. (1998).

Toward this end we ask senders in the trust game about the threshold of money returned by the receiver below which they would feel cheated. We ask this for each possible amount

[1]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.

[2]Berg et al. (1995) is among the top 1% of research items (by number of citations) on RePec and Google scholar lists over 1400 citations.

[3]The response in Berg et al. (1995)—"trust in reciprocity," where reciprocity is defined by a positive return on investment—fares surprisingly poorly at reassuring doubts.

[4]This needs not be a damning critique as, for instance, we can feel cheated by a taxi driver who takes a long route even when we have not agreed beforehand that the shortest route is the goal.

they could send. To avoid this elicitation mechanically affecting their behavior in the game we ask this *after* they have played the trust game but before they know any outcome. Our data suggest that respondents have a clear notion of cheating even when the receiver makes no explicit promise: the vast majority of senders would feel cheated by a negative return on their trust/investment.[5] Interestingly, within this group, a sizable minority of participants—around one third—have a more demanding notion based on an equal split rule: they would feel cheated by a return amount that is less than half of the total money their counterparts control.[6]

The fact that participants in the trust game have a notion of what cheating behavior is does not necessarily prove that trust is affected by it. To move a step in this direction we also elicit senders' beliefs about the proportion of cheaters in the (experimental) population, where cheating is defined according to their notion. We find that beliefs about this proportion—a measure of the subjective probability of being cheated—do have a strong impact on senders' behavior. Those who hold a higher probability of being cheated are more likely to send no money at all. To the best of our knowledge, this is the first paper directly examining the relationship between behavior in the trust game and participants' own notion of cheating.

Once a notion of cheating is obtained we can study how it affects receivers' behavior. First, we find that receivers' estimates of senders' notions of cheating strongly affect how much receivers return. This suggests that receivers consider whether or not senders will feel cheated when deciding how much money to return. Second, since we can define whether receivers decisions entail cheating, we can study what restrains their cheating and who is more prone to cheating. We uncover interesting results: the values participants' parents emphasized during their upbringing affect receivers' cheating decisions. Instilled values of fairness and altruism matter for the highest trust levels (high amounts sent)—where receivers are given the most discretion in determining their pairs' outcomes. For the lowest level of trust (low amounts sent), however, where receivers have minimal discretion and, in particular, always earn less than their counterpart, none of these values has a statistically significant impact.

[5]This is true even for low send amounts, implying the otherwise-puzzling phenomenon of senders effectively demanding high percentage returns on small investments, and low percentage returns on large investments.

[6]Such an equal split rule makes sense since trust game receivers essentially face a dictator game situation, and a widely-accepted norm of behavior in dictator games is to divide money equally.

All together, our experimental results suggest that the trust game unambiguously involves trust once senders' personal normative standards are taken into account. We find that the vast majority of these personal cheating definitions entail (at least) a positive return on investment. Senders' behavior is strongly affected by their definition of cheating. For receivers' behavior, instilled values are significant predictors of cheating behavior for high levels of trust.

Our contribution is related to at least two recent strands of literature. First, it links well with the recent debate about the meaning of trust in the trust game started by Glaeser et. (2000) and expanded by Cox (2004), Bohnet and Zeckhauser (2004), Toldra et. al (2007) and Fehr (2009). The focus of these papers is on what trust behavior in the game measures, whether preferences or beliefs; our focus is different and concerns whether observed trust behavior actually reflects people's expectations of cheating even when what is cheating is left unspecified.

Second, our results contribute to the more general 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. Focusing on the relationship between receivers' behavior and 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.[7] Our paper however complements and extends Gneezy's analysis in two ways. First, we show that the moral forces at work in Gneezy's data operate outside of the context of deception as there is no communication nor unambiguous moral standards in our trust game. We still find that receivers' cheating decisions are affected by the extent of their counterparts' vulnerability as in Gneezy (one could have expected this extra "moral wiggle room" to lessen the impact of morality on behavior). Our analysis also moves a step forward to inquire why people act this way and identifies the reason as the values instilled by one's parents. Finally, our results are consistent with Charness and Dufwenberg (2006) where aversion to guilt constrains cheating in binary trust games. While they examine guilt through the analysis of second order beliefs—receivers' beliefs about senders beliefs about what receivers' *will*

[7]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.

do—we examine how participants' first-order beliefs about others' personal definitions of cheating affect behavior. While we collect both first- and second-order beliefs, we find the effect of first-order beliefs to be consistently larger than second-order beliefs.[8]

The remainder of the paper proceeds as follows: Section 2 details the experimental design; Section 3 presents the results; Section 4 puts them in perspective; the final section provides concluding remarks.

2 Experimental Design

The study involved a total of 122 participants, recruited at LUISS Guido Carli University in Rome, Italy. Sessions were conducted on-line to provide anonymity between subjects and to minimize experimenter/demand effects.

Participants played a mostly-standard trust game (instructions in the appendix). The game involves two roles: sender and receiver. Each sender is endowed with 10.5 euros and chooses whether to keep this endowment, or send a positive amount to their randomly-chosen anonymous receiver. Sending a positive amount costs 0.50 euros. Upon paying this fee, a sender can send any positive integer amount to the receiver. This amount is increased according to a concave "production" function before reaching the receiver.[9] Finally, the receiver decides to return some, all, or none of this increased amount to the sender.

Participants were randomly paired and within each pair one participant was assigned the role of sender while the other was assigned the role of receiver. Participants' strategies were collected using the strategy method: before discovering which role they were assigned, participants submitted a complete contingent strategy for each role. Each participant specified how much they would send in the role of senders, and, for each possible amount they could receive, how much they would return in the role of receivers. The order in which

[8] We do not report the results using second-order beliefs for brevity and because we also find that these two measures are highly correlated. Notice that this need not be the case, because second-order beliefs are receivers' beliefs about senders beliefs about how much *will* be returned by receivers, rather than how much *should* be returned by senders. The latter can be thought of as necessarily including a moral component, while this is not a necessary component of the former. The fact that they are related suggests that senders believe receivers *will* act morally. However, we realize that the difference in the estimated magnitudes of the effects of first- and second-order beliefs could be due to our elicitation procedure so we do not read too much into the estimated differences, and for this reason also omit the results on second-order beliefs.

[9] If the sender sends S euros, the receiver receives $8S^{0.5}$ euros. 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 production function was presented to participants as a table to facilitate comprehension.

participants submitted their strategies—whether first for sender, then for receiver or first for receiver and then for sender—was randomized. Additionally, 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 screens asked only one question: "if the sender sends S euros and you therefore receive $f(S)$ euros, how much will you return?" For each separate screen, S was replaced with exactly one of the 10 possible amounts a sender could send ($S \in \{1, \ldots, 10\}$), 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.[10] This maintains comparability across observations without inducing any undue consistency in receiver strategies.

To investigate the role played by cheating in the trust game, we adopt a broad definition of trust constructed from an interdisciplinary review of the trust literature by Rousseau, et al. (1998):

> "Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of others."

The key phrase, "positive expectations," indicates that the presence of trust depends on individuals' conceptions of the normative standard by which behavior is judged. In this light, the normative standard suggested in Berg et al. (1995) would be a weakly positive return on investment[11].

Based on this, participants in our experiment were asked the following question:

> "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 [x] euros and B were to therefore receive [$f(x)$] euros, you would need back how many euros?"

[10] The order used was $S = 7, 4, 8, 3, 9, 10, 2, 1, 6, 5$.

[11] Other standards are obviously possible. Equally obvious is the fact that normative standards can be well-defined in the absence of explicit contracts. If contracting is possible, on the other hand, it can be codified as a minimal acceptable outcome: whenever counterparties' actions fall below this standard, individuals can feel cheated or betrayed. Note also that normative standards can (and do) vary across individuals, potentially obfuscating the role of trust in observed trust game behavior. For example, our own values can color our expectations of others' values and behavior in both a normative and statistical sense, through, e.g., false consensus (Ross, Greene and House, 1977).

There were 10 of these questions in total, one for each positive amount a sender could send. In each question $[x]$ and $[f(x)]$ were replaced by the appropriate numbers, and "[sender]" and "[receiver]" did not appear.

Next, but still before knowing which role they were assigned, participants discovered there would be a belief-elicitation portion of the experiment. They were told that in this section they would be asked to estimate other participants' responses and actions, and that they could earn additional money according to the accuracy of their estimates[12]. Each participant was asked to estimate the average of others participants' strategies in the trust game as well as the average of others' responses to the cheating questions above. Participants also estimated the proportion of other participants who would not cheat them according to the participants' own definitions of cheating. This provides us with a measure of the subjective probability of not being cheated when playing as a sender. These questions were asked *after* participants submitted their complete contingent strategies, but *before* knowing their assigned roles. This way we avoid mechanical correlations between behavior in the trust game and their notion of cheating and probability of being cheated.

Finally, roles were revealed and earnings from the trust game were determined combining, within each pair, the sender's strategy with the receiver's strategy. Participants were also informed which randomly-chosen estimation would count toward their potential earnings and how much this estimate earned them. Ten percent of participant pairs were randomly chosen to be paid their potential earnings.

Complementing the experimental results, for each subject we have data from a previously conducted survey containing basic demographic information and a measure of risk aversion. Risk aversion in the survey was elicited using the procedure in Holt and Laury (2002).[13] There was a considerable time lag (from 20 to 60 days) so survey responses are unlikely to have affected trust game behavior directly. On the other hand, this temporal distance was small enough so that stable traits, such as risk aversion or instilled values, likely did not

[12]The exact instructions appear in the appendix. To provide the appropriate incentives, subjects were paid according to a randomized quadratic scoring rule (Schlag and van der Weele, 2009). This scoring rule has been shown to be theoretically incentive compatible even when subjects are not risk-neutral.Estimates were remunerated according to their accuracy, with perfect estimates paying 5 euros. At the end of the experiment, one estimate for each participant was chosen to count toward their earnings.

[13]Briefly, this procedure asks participants to make a sequence of ten choices, each of which involves a choice between a relatively risky lottery and a relatively safe lottery. The sequence is constructed so that more risk averse individuals will switch from preferring the safer lottery to the riskier lottery later in the sequence. Therefore, this measure is increasing in risk aversion.

change in the meantime.

3 Results

We establish three main results: *i)* we uncover the notion of cheating in the trust game; *ii)* we show that beliefs about cheating are relevant in senders' decisions to trust; *iii)* we show that these cheating notions also affect receivers' behavior, and moreover, we investigate possible determinants of receivers' cheating decisions. The descriptive statistics for all the variables in the experiment are reported in Table 1.

3.1 Notions of Cheating

We start by looking at the notions of cheating as reported by senders (Table 2A). The vast majority of senders report they would feel cheated if their co-players do not return at least as much is sent. This is true for any amount a sender could possibly send. Depending on the amount sent, the fraction of senders who feel cheated by a negative return on their investment ranges from 93 percent down to 76 percent with higher fractions for smaller amounts sent.[14]

As a second step we plot kernel density estimates for individuals' own cheating definitions and their estimates of other participants' notion of cheating considering sending 1 euro (Figure 1). It is apparent from the figure that participants' own cheating notions and their estimates of others' cheating notions roughly cluster around 1 euro and 4.025 euros (indicated by two vertical bars): the first value implies cheating defined by negative return on investment, while the latter implies cheating defined with respect to an equal split of the money receivers get from the senders. The latter cheating notion may seem extreme, but fits with the interpretation generally given of the trust game that receivers are essentially in the position to divide a pie[15].

Figure 2 presents the analogous kernel density plots for the rest of the possible send amounts. There are two points to notice. First, cheating definitions continue to roughly cluster around the two norms of equal-splits and positive return on investment across send

[14] In table 2A 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 notion. An alternative would be to define cheating as *amount returned* - (*amount sent*$+0.50) < 0$. Results if this alternative criteria is used are unchanged but all proportions shown in Table 2A are a bit scaled down.

[15] It is well-known that a common decision rule when dividing a fixed sum is to divide the sum into equal shares (c.f., evidence on the Dictator Game in Camerer, 2003)

amounts. Second, estimates of others' cheating definitions follow the same patterns. Thus, not only do participants adopt these two definitions of cheating, but they realize others do as well. As the distance between the values defining these two notions of cheating decreases—which happens as amounts sent increase—both the distribution of beliefs about others' cheating notions and the distribution of participants' own cheating notions become more uni-modal.

From the analysis of the kernel densities it seems that the implied notions of cheating fall roughly into two categories. While a (weakly) positive return on investment is enough to characterize the vast majority of participants' notion of cheating, a large minority—roughly one third—demand half of the total money receivers receive in order to not feel cheated. We label these participants "equal splitters". This can also be seen from the first column of Table 2B, where we report the overall proportion of equal splitters for each send amount: this proportion is fairly constant and always roughly equal to one third. The next two columns show that the proportion of equal splitters does not differ if we distinguish between men and women, implying no gender-driven differences in the notion of cheating.

To see whether there is individual consistency with respect to the cheating definition, we restrict the sample to those who demand an equal split when sending one euro[16] (this includes 37 equal splitters out of 122 participants) and plot their kernel density estimates of cheating notion. Participants who define cheating in terms of an equal-split when sending 1 euro tend to use this cheating definition for other possible send amounts, and that they tend to believe others define cheating in the same way (Figure 3).

To sum up, people in the trust game do have a very clear notion of cheating even though no promise is made on the side of receivers. Does expected cheating affect senders' behavior? Based on these notions, do receivers actually cheat and what drives their decisions to cheat? We now turn to addressing these questions.

3.2 The Effects of Cheating Beliefs on Senders' Behavioral Trust

Our main result is that participants' decision about whether to trust or not depends crucially on how likely they think it is that they will be cheated according to their personal notions of being cheated. To show this we construct a measure of an individual's beliefs about the

[16]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 minimum amount of between 4 and 5, inclusive, in order to not feel cheated.

proportion of non-cheaters in the (experimental) population by averaging his or her answers to the following set of 10 questions ($x = 1, 2, \ldots, 10$):

> "If you send €x and B therefore receives €$[f(x)]$, what percent of B's will return enough money so that you don't feel cheated?"

The resulting measure of beliefs about population trustworthiness theoretically ranges from 0 to 1, with 1 indicating no receivers will cheat (all are trustworthy) and 0 indicating all receivers will cheat (none is trustworthy). We call this measure the subjective probability of being cheated. Figure 4 plots the kernel density of this probability and documents a modal value at around 0.5 (not far from the fraction of cheaters in the pool - see Table 1). Additionally, we construct a measure of senders' beliefs about the proportion of the money they send that will be returned by averaging their 10 return proportion estimates. The resulting averages range from a low of 0.08 to a high of 3.02 in the data. This is a measure of senders expected (gross) return.

We then estimate a probit model of senders' (binary) trust decisions—whether or not to send a strictly positive amount of money to their co-player—where the dependent variable takes the value of 1 if the sender trusts and 0 otherwise. The resulting estimates are shown in Table 3A. They imply that the probability of being cheated plays a significant role in the decision to trust even when controlling for expected pecuniary returns. The estimated effect is large: in the simplest specification (column 1), the estimated coefficient implies that going from a belief of 0 (sure probability of being cheated) to a belief of 1 (sure probability that receiver will not cheat) raises the chances of trusting by 64 percentage points, almost as much as the sample fraction of trusters.[17] Increasing the probability of not being cheated by one standard deviation raises the probability of trusting by as much as 15 percentage points—or 20% of the fraction of people trusting. In this first specification the gross return does not play a relevant role in the decision to trust once the probability of being cheated is controlled for. In column 2, we look at the interaction between expected pecuniary returns from trusting and expected betrayal. There are two points to notice: first, returns have now a direct positive and significant effect on the decision to trust; second, they also affect the sender's decision indirectly by changing the sensitivity to betrayal: for low expected profits,

[17] Since the game involves a fixed cost from trusting (50 cents) it may be desirable to send nothing even if the sender is sure that the receiver does not cheat. This explains why some people may send nothing even when the probability of not being cheated is 1; but they are few.

expected betrayal is quite important in the decision to trust but it matters much less as expected pecuniary profits increase. A one standard deviation increase in pecuniary return from the sample mean lowers the sensitivity to the probability of (not) being cheated by 37%. This suggests that money can be powerful mechanism in luring people into trusting by weakening their fear of betrayal.

Columns 3 and 4 replicate the same specification adding a rich set of controls (we include controls for gender, age, math ability and risk aversion collected from a separate, previous, survey,[18] and dummies for family income, in addition to those, in column 4). Our results are robust to adding these controls which, except for income have little predictive power.

In Table 3B we run a Heckman selection model for the amounts sent; to achieve identification we exclude income from the send amount and include it on the probit decision to trust. The justification for this is that the fixed cost of sending—the 0.50 euro fee—matters most for relatively low income senders than for high-income ones, giving us some heterogeneity in the incentives to send. With this assumption, we find that the amount sent is positively and significantly affected by the size of pecuniary returns. Most interestingly, it is also significantly affected by the probability of not being cheated by the receiver. Increasing the latter by one standard deviation increases the amount sent by 12.6% of the sample mean (among the active senders). We conclude that beliefs about risk of betrayal matter both for the choice of trusting at all as well as for the amount of trust as measured by amounts sent. The latter are also affected by senders' aversion to risk.

3.3 What Drives Receivers' Decision to Cheat ?

If the belief about (absence of) cheating drives the decision to trust (and how much), then the question of what drives cheating becomes important for all the same reasons that trust itself is important. We investigate the behavior of receivers along two dimensions: first, we study whether receivers incorporate the senders notion of cheating in their choice of how much to return. That is, we investigate whether receivers' decisions about how much to return is constrained by what they believe the senders' definitions of cheating are. Second, we study what drives receivers' decisions to cheat and whether instilled values restrain

[18]This survey was conducted well beforehand: the time lag between completing the survey and taking part in the experiment ranges from 20 to 60 days. We measure risk aversion through a Holt and Laury (2007) procedure which results in an index from 0 to 10, with higher index values indicating higher risk aversion. The procedure was conducted in an incentive compatible manner.

them from doing so. We can address this latter question directly because we know when receivers cheat according to their own estimates of others' cheating definitions. Additionally, we have information from the previously-conducted survey about the values participants' parents emphasized during their upbringing.

We examine the first issue in Table 4 panel A which presents OLS estimates of the amounts receivers return as a function of their estimates of others' cheating definitions and the standard demographic controls. We run a separate regression for each amount a sender could send resulting in 10 separate estimates. For all possible amounts senders could send, receivers' estimates of senders' definitions of being cheated plays a statistically and behaviorally significant role in the amount receivers plan to return. The estimates suggest that for each additional euro receivers believe senders need back in order to not feel cheated, return amounts increase from a minimum of an additional 35 cents (if sent 2 euros) to a maximum of an additional 60 cents (if sent 1 euro).[19]

To investigate the second issue—the drivers of cheating decisions—we construct a dummy variable taking the value of 1 whenever a receiver returns less than the minimum they believe others need back in order to not feel cheated and 0 otherwise, for each amount a receiver could send. In other words, we construct a dummy indicating when receivers *intentionally* cheat. We then relate this variable to two sets of observables: receivers' demographic characteristics and various measures of instilled values. Table 5 presents our estimates of receivers' propensities to intentionally cheat for three send amounts: $1, 5$ and 10 euros. We choose these send amounts because they represent three fundamentally different situations. If the sender sends 1 euro, then the receiver always makes less money than the sender since the sender retains 9 euros while the receiver only receives 8.05 euros.[20] If the sender sends his or her entire endowment, the receiver has full discretion to decide earnings for both parties. Finally, sending 5 euros is an intermediate situation. Receivers can choose to make either more or less money than senders, but they cannot completely determine both parties' earnings.

Among the demographics, the most relevant result is a strong effect of gender on cheating propensity. Males are systematically more likely to cheat than females: at sample means,

[19]Differences in impact do not seem to have a systematic pattern across send amounts; a formal test does not reject the null hypothesis that the effect of others' cheating definition on amount returned is the same for all amount senders send.

[20]Recall that the receiver began with no endowment, so he or she is truly behind in terms of earnings in this situation.

being male raises the probability of cheating the sender by 18 percentage points (31% of sample mean) when the send amount is 1 and by 16 percentage points (42% of sample mean) when the send amount is 10. Interestingly, the gender difference cannot be attributed to differences in moral standards since we are controlling for them. One interpretation is that men are more greedy and selfish than women and thus more willing to exploit opportunities to expropriate money. This view has received support in experimental and survey-based studies showing that men are more self-oriented than women (for instance Eagly and Crowley (1986); Eckel and Grossman (1998); Reiss and Mitra (1998)). Lending further support to this view, notice that in Table 5 the gender effect on cheating is stronger when amounts sent are larger.

Second, we add as an explanatory variable receivers' own notions of cheating. Controlling for their expectations about others' notions of cheating (which by construction has a positive effect on the probability of cheating) receivers that have higher standards—i.e. would feel cheated unless they were given back a lot when playing as senders—tend to be less likely to cheat at all levels sent. 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 from deviating from it (and doing to others what you would not want them do to you) is thus stronger. Consistent with this we find that the effect of the own notion of cheating is much stronger at low levels of amount send and weaker at high levels: the marginal effect of an increase in the own notion of cheating at send amount 10 is half that at send amount 1.

Turning to the effect of instilled values on the choice to cheat, the results suggest that receivers' instilled values have a significant impact in the latter two situations, and that receivers treat these situations differently. When they have full discretion (senders send 10 euros), values of fairness (fair share) and altruism (help others) reduce the likelihood of cheating. When receivers have considerable, but not full, discretion (senders send 5 euros), civic values such as loyalty and acting so as to induce good in others govern the cheating decision. When receivers always do poorly (send 1 euro), values play no role in contrasting the cheating decision. These results imply that what instilled values do is to constrain the receiver's incentive to cheat when the act of cheating may cause serious damage to the sender. When senders send very little, cheating by the receiver does not harm them much

and there is thus little scope for the norm; when senders send a lot cheating can harm them considerably justifying the intervention of the norm to discourage this behavior. This is consistent with Gneezy's (2005) finding that people are less likely to deceive when the harm that deception imposes to the deceived is large relative to the benefit of deception for the deceiver. Our results trace this behavior back to the cultural norms that society endows us with and show that these norms can be quite powerful: at sent amount 10 removing altogether norms of fairness and altruism would result in an increase in the probability of cheating by 42 percentage points.

4 Discussion and interpretation

To put these results in perspective and shed more light on what sort of preferences can explain the receivers' cheating decisions, Figure 5 plots the fraction of receivers who cheat at each send amount after partialling out the effect of the expected notion of cheating. The share is 46% at send amount 1 and drops continuously to 9.3% at send amount 10. Obviously this pattern is inconsistent with a scheme where receivers act selfishly looking only at their own monetary payoff. In this case all would cheat and the pattern would be flat at a share equal to 1. However it is also inconsistent with models where receivers face a fixed cost of cheating: since potential pecuniary gains from cheating increase in the amount sent, such a model would predict a non-decreasing relationship between amount sent and cheating propensity.[21] Our results also appear to be inconsistent with the most widely-cited social preferences models—inequality aversion (Fehr and Schmidt, 1999) and social welfare preferences (Charness and Rabin, 2002). Since receivers actions do not create surplus, but only distribute it, according to both of these models receivers should never allocate less money to themselves than to their co-players. However, the majority of receivers choose to put themselves strictly behind their counterpart when the amounts sent are small. This proportion declines monotonically going overall from a fraction of 84% when the amount sent is 1 to 44.2 when the amount sent is 4 to 18% when the amount sent is 10.[22]

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

[22] It should be noted that this is not out of necessity, except when the amount sent is 1. But even there, the figure of 84% includes only those receivers *returning* a strictly positive amount, so these receivers are willingly putting themselves even further behind their counterpart, thereby increasing inequality, without any offsetting increase in surplus or personal gain.

A simple model that accounts for the estimation results in Tables 4 and 5 ond for the cheating pattern in Figure 5 incorporates a widespread and intuitive moral standard: irrespective of what constitutes cheating, it is generally viewed as "more wrong" to cheat the vulnerable. For example, crimes against the elderly and the very young are viewed as particularly reprehensible regardless of the crime. Well-established moral codes such as that found in the Bible instruct us to take special care of those likely to be vulnerable (e.g., foreigners). This is the point made by Gneezy (2005). In the context of the trust game, sending higher amounts makes senders more vulnerable and consequently makes cheating them morally and psychically more costly.

To model this, we add a moral cost function to receivers' standard pecuniary utility function. The moral cost function has three arguments: the vulnerability of the sender as measured by the amount sent, s, a fixed cost of cheating and a term measuring the degree with which the receiver cheats and defined by the distance between the receiver's estimate of the sender's cheating notion and the amount the receiver returns.

Denote this extra moral cost function $m(s, K_j, dist(r, c_j(s)))$. Receiver's utility is given by:

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

In 1, $f(s)$ is 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 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:

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

In equation 2, $u(f(s) - r)$ is, again, the receiver's standard pecuniary utility over money from receiving $f(s)$ and returning r. Assume that u 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 is captured by $v(s)$ which we assume is increasing. Notice that setting $\alpha_j = 0$ reduces receivers' utility to standard amoral preferences and that if $\alpha_j > 0$, setting $\gamma_j = v(s) \equiv 0$ implies receivers have simple fixed-cost-of-cheating preferences.

This specification for receiver utility can explain: a) the dependence of the decision to cheat on meaures of morality and of the cheating notion, in a way that is consistent with the results in Table 5; b) the positive correlation between amounts returned and the notion of cheating shown in the Table 4, panel A; 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.[23]

This simple preference specification can also account for another feature of the data:

[23]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, s^*, is then 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.

conditional on cheating people do not go as far as returning nothing but they send some money back. The amount returned should depend on the expected notion of cheating (people should return more if they expect the others require more money back in order not to feel cheated), 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: hence the amount returned should move one to one with the expected cheating notion. The latter prediction is also shared by the fixed cost of cheating model; the first is not and is specific to this model.

Table 4, panel B puts these predictions to a test. We split the sample between cheaters and non-cheaters and report regressions for the amounts returned, while adjusting for selection. The two exclusion restrictions for the first stage probit (not reported for brevity) are based on the findings of Table 5: the own notion of cheating and a measure of the intensity of the norm of fairness. We interpret them as capturing differences across individuals in the fixed cost of cheating which, while affecting the decision to cheat or not to cheat, has no impact on how much to return. Consistently with Table 5, we report regressions for three amounts sent: 1, 5 and 10 euros. 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 more sensitive to expected cheating notions for non-cheaters than for cheaters. In fact, for non-cheaters the estimated coefficient on expected cheating notions is not significantly different from 1 in any case, while for cheaters this coefficient is significantly less than 1 in two out of the three cases.

5 Concluding Remarks

There has been much debate over whether behavior in the trust game has anything to do with trust. In particular, it has been argued that, since no explicit promises or contracts are involved in the trust game, cheating is not defined. This experiment addresses this critique by eliciting participants definitions of being cheated as well as their beliefs about others' definitions of being cheated. We find that the trust game invokes two intuitive notions of being cheated: a negative return on the amount sent and a less-than-equal split criteria. Using participants' personal cheating definitions, we find that beliefs about the propensity of others to abstain from cheating—their trustworthiness—play a significant role in senders'

decisions in the trust game.

Moreover, collecting information on cheating notions and values instilled by parents allows us to investigate what drives cheating. We find that the values instilled by parents play heavily into cheating decisions, and that the impact of values is magnified by counterparty vulnerability. This evidence complements and extends the growing literature on how the interaction between moral concerns and pecuniary preferences affects behavior.

References

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

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

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

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

[5] Fehr, Ernst (2009), "On the Economics and Biology of Trust", Presidential Address, European Economic Association, Journal of the European Economic Association, forthcoming

[6] 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.

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

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

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

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

[11] 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.

[12] 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.

[13] 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.

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

Table 1: Descriptive Statistics

	Mean	Std Dev	Min	Max	N
Male	0.42	0.495	0	1	120
Age	22.38	2.051	19	29	120
Math score	7.77	1.18	4	10	117
Inc<30K	0.16	0.368	0	1	113
30≤Inc<45	0.16	0.368	0	1	113
45≤Inc<70	0.28	0.453	0	1	113
70≤Inc<120	0.25	0.434	0	1	113
Inc≥120K	0.15	0.359	0	1	113
Risk aversion	4.77	2.095	1	10	117
Send decision (binary)	0.73	0.446	0	1	122
Send amount	3.93	3.315	0	10	122
Average return proportion	1.31	0.669	0	3.93	122
Average expected return proportion	1.29	0.578	0.09	3.02	122
Instilled value: induce good	6.18	2.570	0	10	119
Instilled value: help others	8.15	1.676	2	10	119
Instilled value: loyalty	6.20	2.607	0	10	117
Instilled value: fair share	7.27	2.214	0	10	118
Risk aversion	4.77	2.095	1	10	117
Expected return from trusting	1.29	0.578	0.09	3.02	122
Expected probability of not being cheated	0.42	0.225	0.02	1	122
Fraction of cheaters	0.46	0.499	0	1	122

Table 2A: Proportion of people who would feel cheated if return is negative

€ Sent	Mean	Std Err	N
1	0.93	0.024	122
2	0.89	0.028	122
3	0.92	0.025	122
4	0.90	0.027	122
5	0.87	0.031	122
6	0.84	0.034	122
7	0.79	0.037	122
8	0.77	0.038	122
9	0.78	0.038	122
10	0.76	0.039	122

Notes: [1] Since the strategy method was used to collect participants' decisions, each row includes observations from all participants rather than only half.

Table 2B: Proportion of people who would feel cheated if return is negative

€ Sent	Overall proportion equal-splitters when sent €x.	Male: Overall proportion equal-splitters when sent €x	Female: Overall proportion equal-splitters when sent €x	Equality of proportions test significance
1	0.30	0.26	0.31	p=0.519
2	0.33	0.34	0.31	p=0.767
3	0.34	0.34	0.31	p=0.767
4	0.37	0.24	0.27	p=0.698
5	0.30	0.26	0.30	p=0.632
6	0.34	0.32	0.33	p=0.921
7	0.33	0.30	0.33	p=0.740
8	0.26	0.22	0.27	p=0.521
9	0.29	0.28	0.29	p=0.945
10	0.33	0.36	0.30	p=0.489

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 their receive. 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 3

Panel A. Cheating expectations and the sender decision to trust

	(1)	(2)	(3)	(4)
Expected probability of not being cheated	0.642***	1.204***	1.172***	1.063***
	(0.207)	(0.279)	(0.335)	(0.311)
Expected return from trusting	0.0180	0.173*	0.178***	0.119***
(expected proportion of money returned)	(0.0426)	(0.0935)	(0.0577)	(0.0432)
Probability of not being cheated x		-0.434***	-0.398***	-0.296***
x Expected return from trusting		(0.158)	(0.121)	(0.0744)
Male			0.0131	0.0364
			(0.0561)	(0.0385)
Age			0.0393	0.0379
			(0.0349)	(0.0389)
Math score			0.0185	0.0363*
			(0.0186)	(0.0219)
Risk aversion			0.00970	0.0215
			(0.0307)	(0.0296)
30≤ Income <45				0.176***
				(0.0360)
45≤ Income <70				0.0288
				(0.0606)
70≤ Income <120				0.0371
				(0.0591)
Income ≥120				0.267***
				(0.0488)
Observations	122	122	114	108

Panel B: Amounts sent, accounting for censoring (Heckman selection model)

	(1)
Expected probability of not being cheated	2.956***
	(1.116)
Expected return from trusting	0.510***
(expected proportion of money returned)	(0.196)
Male	-0.0946
	(0.202)
Age	0.392*
	(0.233)
Math score	0.180
	(0.240)
Risk aversion	-0.213**
	(0.102)
Constant	-6.942
	(7.814)
Observations	108

Notes: [1] Robust standard errors, clustered by session, in parentheses. *** = significant at 1%, ** = significant at 5%, * = significant at 10%. [2] Each column refers to marginal effects estimates from a Probit model. [3] Dependent variable is whether participant sends money in role of sender (binary). [4] "Probability of not being cheated" is participants' average estimate of the proportion of receivers who will not cheat the participant, according the participant's own definition of being cheated. The average is taken with respect to the 10 measures in the data for each participant: proportion of receivers who will not cheat conditional on sending 1 euro, …, proportion of receivers who will not cheat conditional on sending 10 euros. [5] "Expected return from trusting" is the participant's estimate of the proportion of money sent receivers will return, averaged over all 10 possible send amounts. [6] Income variables refer to (self-reported) annual family income from all sources, in thousands of euros, net of taxes. Excluded category is "below 30 thousand euros". [7] The dependent variable in the Heckman model is the amount senders chose to send, the exclusion restriction in the selection equation is a set of family income dummies.

Table 4: Senders expected cheating notion of others and the amounts they return

A. Amounts returned and cheating notion

	Return amount when sender sends i euros, i=1,…,10.									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Expected	0.642**	0.356**	0.506**	0.340***	0.360***	0.421***	0.453**	0.385**	0.461*	0.458*
Cheated notion	(0.183)	(0.0838)	(0.129)	(0.0478)	(0.0423)	(0.0647)	(0.124)	(0.0914)	(0.162)	(0.158)
Male	-0.282	-0.257	0.0877	0.0175	-0.655**	-0.516	0.430	-0.654	-0.769	-0.892
	(0.210)	(0.246)	(0.584)	(0.320)	(0.139)	(0.572)	(0.484)	(0.517)	(0.826)	(1.162)
Age	0.0150	0.167*	0.128	0.155	0.120	0.0916	0.128	0.170	0.334	-0.0629
	(0.0243)	(0.0681)	(0.144)	(0.151)	(0.0754)	(0.0902)	(0.153)	(0.196)	(0.177)	(0.138)
Math score	0.270*	0.329**	0.196	0.101	0.353*	0.526**	0.459	0.533	0.733**	0.350
	(0.110)	(0.0830)	(0.155)	(0.129)	(0.117)	(0.154)	(0.237)	(0.278)	(0.190)	(0.181)
30≤ Income <45	-0.658	-0.0384	-0.492	1.661*	0.816*	1.096*	-0.131	3.027*	2.006	1.254
	(0.405)	(0.457)	(0.566)	(0.683)	(0.320)	(0.420)	(0.709)	(1.222)	(0.911)	(0.811)
45≤ Income <70	0.0896	-0.431**	-0.562	0.179	0.00978	0.0895	-1.555**	0.586	0.594	0.765
	(0.346)	(0.0752)	(0.622)	(0.322)	(0.180)	(0.445)	(0.352)	(0.342)	(0.500)	(0.742)
70≤ Income <120	-0.291	-0.558	-0.663	0.0175	-0.109	-0.455	-1.198**	0.332	-0.282	-0.685
	(0.351)	(0.303)	(0.499)	(0.300)	(0.626)	(0.567)	(0.302)	(0.540)	(0.679)	(0.462)
Income ≥120	0.389	0.405	0.201	1.226	2.020***	1.770	1.268	3.884**	3.434*	3.910
	(0.509)	(0.404)	(0.232)	(0.946)	(0.204)	(1.160)	(0.564)	(0.793)	(1.323)	(1.815)
Constant	-1.714	-4.198*	-2.020	-1.395	-1.975	-2.992	-1.893	-4.248	-9.446	2.585
	(0.948)	(1.696)	(4.599)	(4.758)	(2.089)	(3.197)	(3.894)	(6.652)	(5.613)	(5.943)
Observations	111	111	111	111	111	111	111	111	110	110
R-squared	0.253	0.126	0.137	0.142	0.184	0.204	0.190	0.253	0.251	0.207

B: Amounts returned by cheaters and by non-cheaters

	Amount sent=1		Amount sent=5		Amount sent=10	
	Cheaters	Non-cheaters	Cheaters	Non-cheaters	Cheaters	Non-cheaters
Expected cheating notion	0.636***	0.812***	0.697***	1.034***	0.466**	0.820***
	(0.104)	(0.152)	(0.216)	(0.215)	(0.211)	(0.132)
Demographics	yes	yes	yes	yes	yes	yes
Mills ratio	1.210**	0.750	2.692*	-2.565*	-1.639	-0.930
	(0.627)	(0.890)	(1.499)	(1.495)	(2.267)	(1.499)
Constant	-0.425	2.274	-9.000	5.523	-5.671	6.316
	(2.089)	(3.475)	(5.884)	(4.419)	(8.615)	(5.310)
Observations	111	111	111	111	111	111
p-value for effect of cheating notion =1	0.0005	0.216	0.1609	0.875	0.0116	0.339

Notes: [1] Robust standard errors, clustered by session, in parentheses. *** = significant at 1%, ** = significant at 5%, * = significant at 10%. [2] Each column in Panel A presents a simple OLS estimate. The dependent variable in column i is the amount a participant will send back if the sender sends i euros, i=1,…,10. The independent variable in column *i* 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. That is, the independent variable is participants' estimates of others' cheating definitions. There is one fewer observation in columns 9 and 10 because one participant failed to submit cheating estimates for (only) these two possible send amounts. [3] The dependent variable in the Heckman model is the amount returned for cheaters and non-cheaters, the exclusion restrictions in the selection equation are given by the own estimate of cheating and the value of fairness as defined in the text.

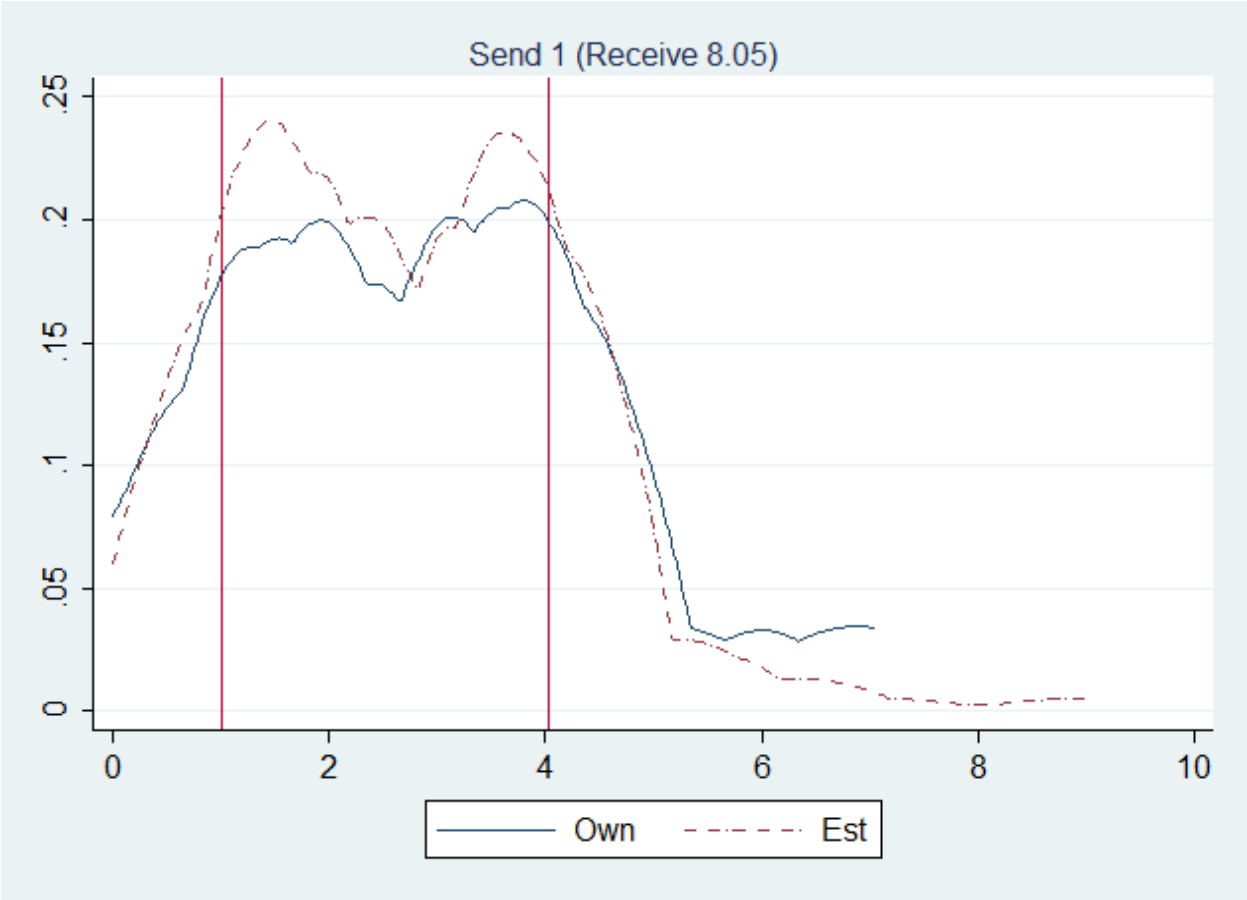
Table 5: Cheating by instilled values

	If sent €1, received €8.05				If sent €5, received €17.90				If sent €10, received €25.30			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Values												
Induce good	-0.000				-0.031**				0.004			
	(0.008)				(0.013)				(0.023)			
Help others		0.017				-0.010				-0.028***		
		(0.013)				(0.025)				(0.010)		
Loyalty			-0.003				-0.043**				-0.013	
			(0.019)				(0.021)				(0.023)	
Fair share				-0.032				-0.034***				-0.078***
				(0.023)				(0.005)				(0.009)
Cheating notion												
Cheating definition, own	-0.117***	-0.138***	-0.107***	-0.118***	-0.090**	-0.080**	-0.084**	-0.088***	-0.057***	-0.054***	-0.060***	-0.064***
	(0.036)	(0.024)	(0.034)	(0.030)	(0.036)	(0.032)	(0.042)	(0.034)	(0.009)	(0.008)	(0.009)	(0.015)
Cheating definition, estimate	0.127***	0.152***	0.129***	0.132***	0.138***	0.127***	0.135***	0.139***	0.090***	0.089***	0.093***	0.109***
	(0.025)	(0.016)	(0.023)	(0.018)	(0.029)	(0.023)	(0.021)	(0.016)	(0.016)	(0.020)	(0.020)	(0.021)
Demographics												
Male	0.166**	0.201**	0.160**	0.182**	0.223***	0.243***	0.278***	0.278***	0.132***	0.140**	0.122**	0.164**
	(0.075)	(0.099)	(0.071)	(0.084)	(0.058)	(0.067)	(0.041)	(0.046)	(0.050)	(0.065)	(0.052)	(0.081)
Age	-0.012	-0.016**	-0.015*	-0.013*	-0.049**	-0.057**	-0.067*	-0.062*	0.009	0.004	0.007	0.015
	(0.008)	(0.008)	(0.009)	(0.007)	(0.025)	(0.028)	(0.036)	(0.034)	(0.015)	(0.018)	(0.017)	(0.019)
Math score	-0.032	-0.056	-0.029	-0.019	-0.036	-0.041*	0.009	0.006	0.026*	0.018*	0.013	0.034
	(0.077)	(0.085)	(0.071)	(0.070)	(0.030)	(0.021)	(0.023)	(0.015)	(0.014)	(0.010)	(0.027)	(0.031)
Risk aversion	0.032*	0.026*	0.033*	0.031*	0.010	0.006	0.017	0.022	-0.014	-0.019	-0.022	-0.030
	(0.018)	(0.016)	(0.019)	(0.016)	(0.015)	(0.013)	(0.021)	(0.021)	(0.024)	(0.026)	(0.024)	(0.028)
30≤ Income <45	0.093	0.087	0.111	0.087	-0.035	-0.046	0.029	0.026	0.082	0.067	0.032	-0.060
	(0.119)	(0.123)	(0.141)	(0.117)	(0.179)	(0.191)	(0.237)	(0.240)	(0.141)	(0.148)	(0.143)	(0.197)
45≤ Income <70	-0.078	-0.127*	-0.057	-0.085	0.155	0.101	0.229	0.186	0.077	0.063	0.051	-0.014
	(0.097)	(0.070)	(0.109)	(0.083)	(0.113)	(0.135)	(0.180)	(0.196)	(0.102)	(0.115)	(0.130)	(0.152)
70≤ Income <120	-0.022	-0.051	-0.034	-0.021	-0.049	-0.054	0.054	0.032	0.065	0.027	-0.026	-0.031
	(0.105)	(0.100)	(0.144)	(0.109)	(0.229)	(0.247)	(0.309)	(0.327)	(0.131)	(0.118)	(0.119)	(0.142)
Income ≥120	0.042	0.025	0.045	-0.008	-0.306*	-0.313*	-0.254	-0.280	-0.226*	-0.250*	-0.235**	-0.350**
	(0.125)	(0.139)	(0.153)	(0.206)	(0.161)	(0.176)	(0.233)	(0.238)	(0.129)	(0.136)	(0.112)	(0.136)
Observations	107	107	105	106	107	107	105	106	106	106	104	105

Notes: [1] Each column refers to marginal effects estimates from a Probit model, with the (binary) dependent variable being "receiver cheats if sent relevant amount." [2] Here cheating is defined by sending back strictly less than the receiver estimates senders need back in order to not feel cheated. This threshold amount is also inserted as a control in each estimate by the variable "Expected cheating notion." [2] Robust standard errors, clustered by session, in parentheses. *** = significant at 1%, ** = significant ay 5%, * = significant at 10%. [3] Math score is individual's self-reported score on required math exams taken during the final year of high school in Italy. [4] 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". [5] "Expected cheating notion" is individual's estimate of the minimum amount senders need back in order to not feel cheated with respect to the relevant sent amount: (1)-(4) with respect to sending 1 euro/receiving 8.05 euros; (5)-(8) with respect to sending 5 euros/receiving 17.90 euros; (9)-(12) with respect to sending 10 euros/receiving 25.30 euros. [6] Observations vary over columns because "don't know" was a possible answer on survey values questions, and observations featuring this non-response have been omitted. Additionally, one participant failed to submit (only) estimates of others' cheating definitions for send amounts 9 and 10. [7] The first four explanatory variables are how much emphasis participants' parents placed on various values, on a 0-10 scale. The specific wording of these normative values was (labels in parentheses did not appear in the text participants saw): (Induce good) "Act so as to induce good behavior in others (e.g. tell someone who is littering he should not do so)"; (Loyalty) "Be loyal to the groups and organizations that you belong to."; (Help others) "Help others."; (Fair share) "Alhways be fair: do not take more than your fair share, and always give others their fair share

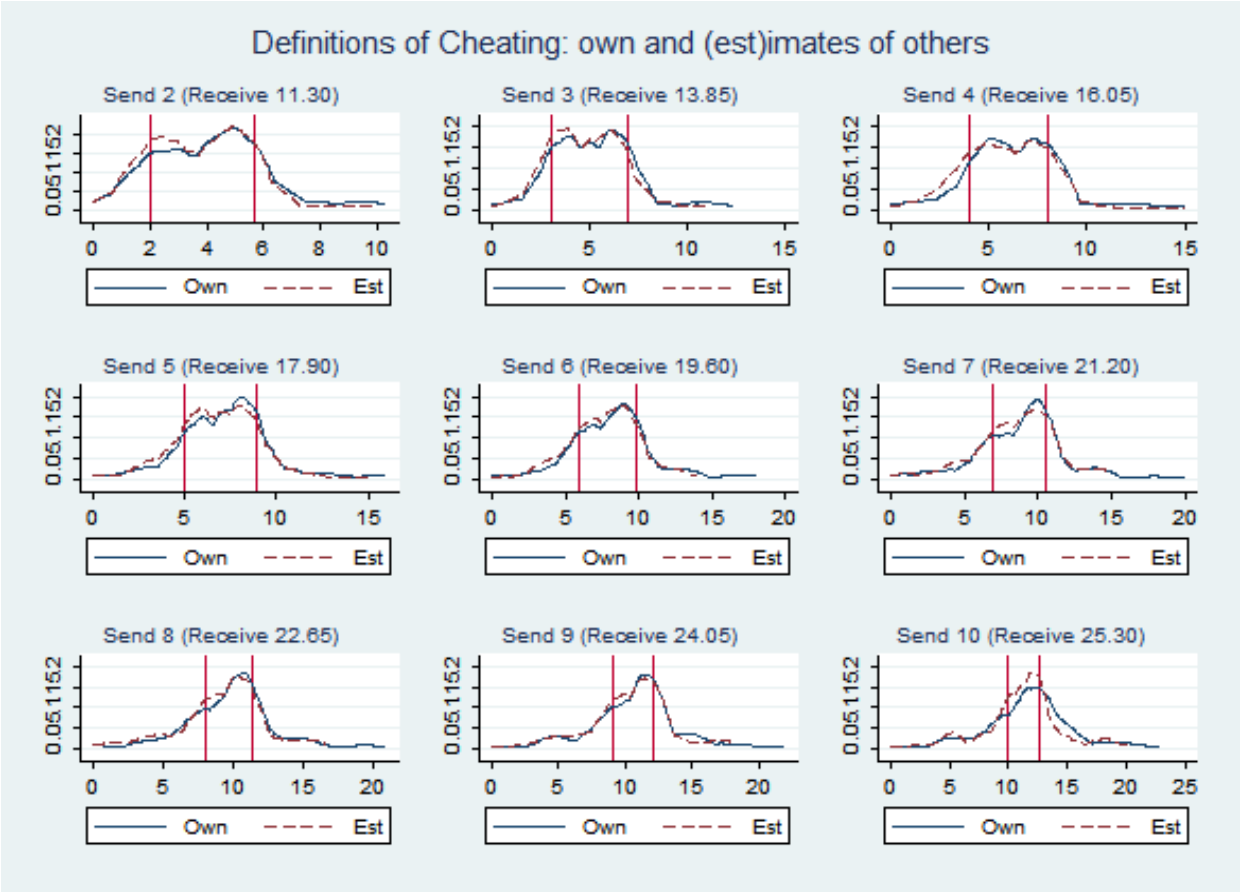
.”

Figure 1: Cheating definitions conditional on sending 1 euro.



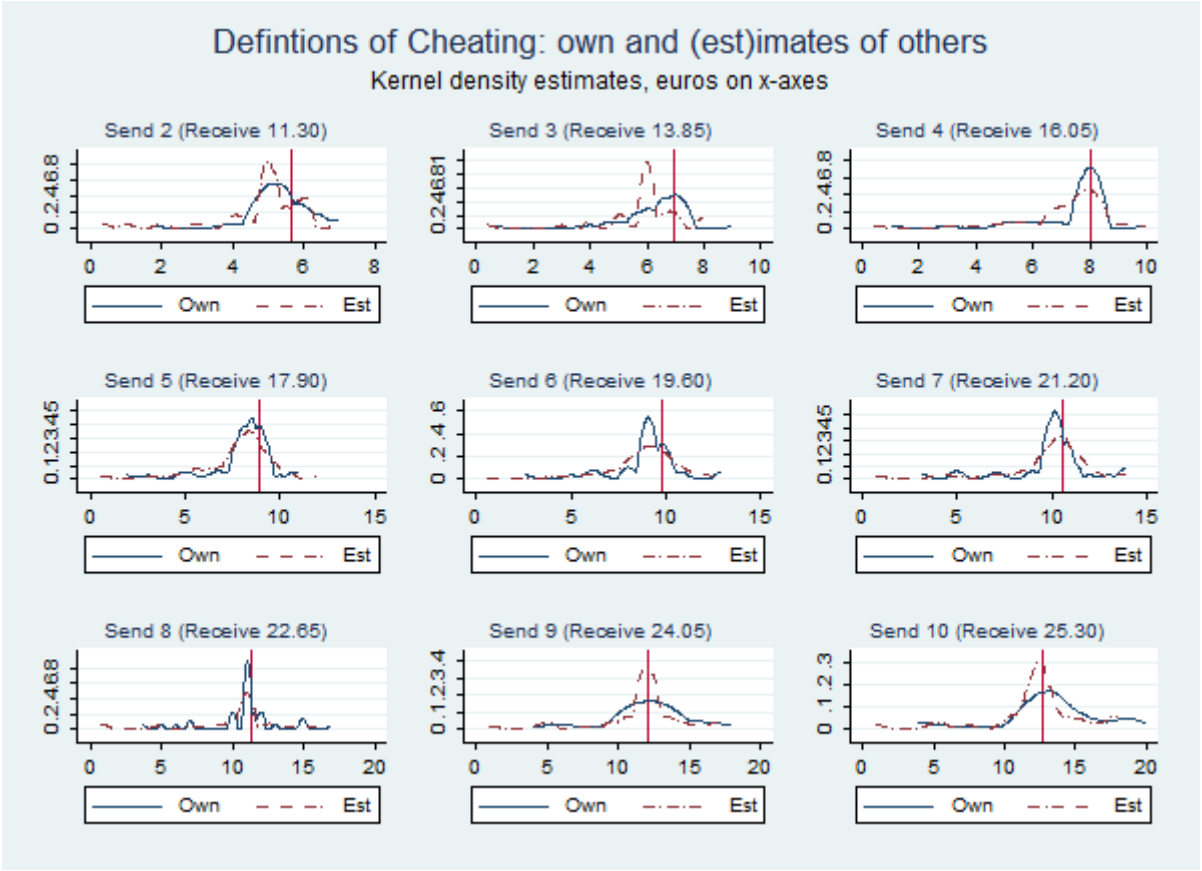
Notes: [1] Kernel Density estimates of participants' definitions of being cheated (Own) and their estimates of others' cheating definitions (Est). [2] Density is on the y-axis, while euros are on the x-axis. [3] The vertical bars correspond to the amount sent (1 euro) and half of the amount receivers receive (4.025 euros).

Figure 2: Cheating definitions by amount sent, own and estimates of others' definitions.



Notes: [1] Kernel Density estimates of participants' definitions of being cheated (Own) and their estimates of others' cheating definitions (Est). [2] For each plot, density is on the y-axis, while euros are on the x-axis. [3] The vertical bars in each figure occur at the amount sent and half of the amount receivers receive, respectively. [4] Notice that as these two values get closer together, the bimodality disappears, indicating one cheating definition related to demanding half of the receiver's money and another related to a positive return on investment.

Figure 3: Consistency in cheating definitions for "Equal Splitters."



Notes: [1] Figure 2 restricts to those defined as "equal-splitters" when sending 1 euro: those who would feel cheated if they receive back less than roughly half of the amount their receiver receives. [2] Because there is a tendency to specify whole-dollar amounts, this definition includes anyone stating they would feel cheated by a return amount less than r, where $4 \leq r \leq 5$. [3] Each figure presents a separate kernel density estimate, with density on the y-axis and euros on the x-axis. [4] The vertical line in each plot represents a value of exactly half the money receivers receive, for each possible amount sent.

Figure 4, Beliefs about the probability of not being cheated

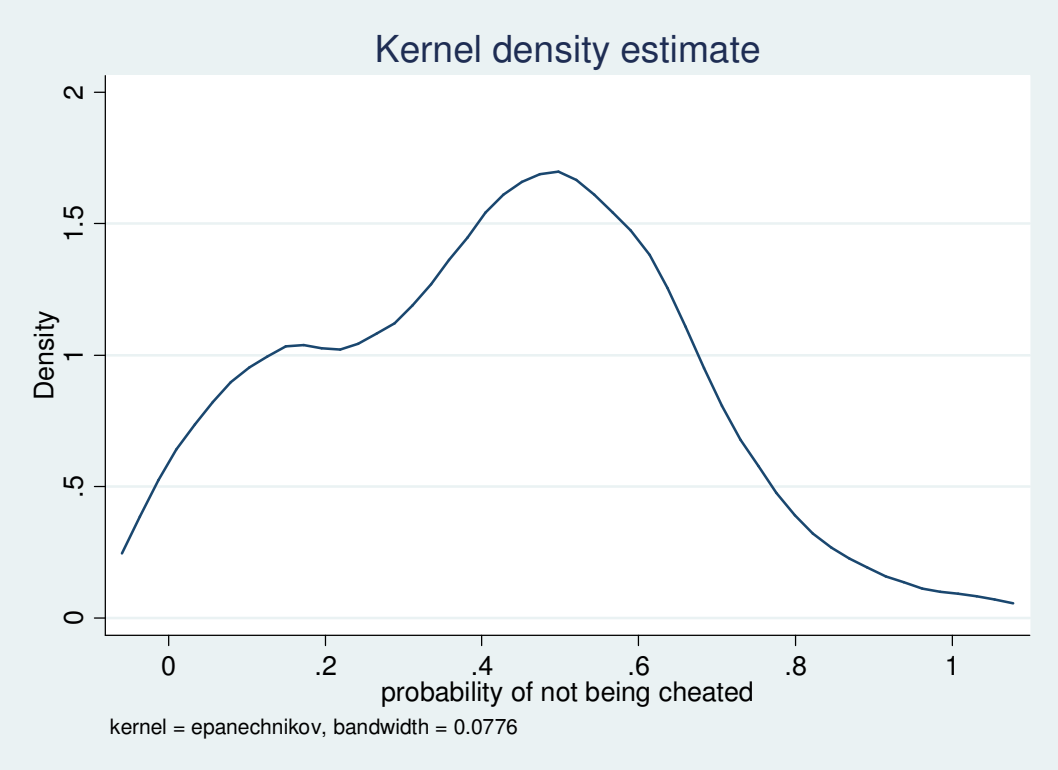
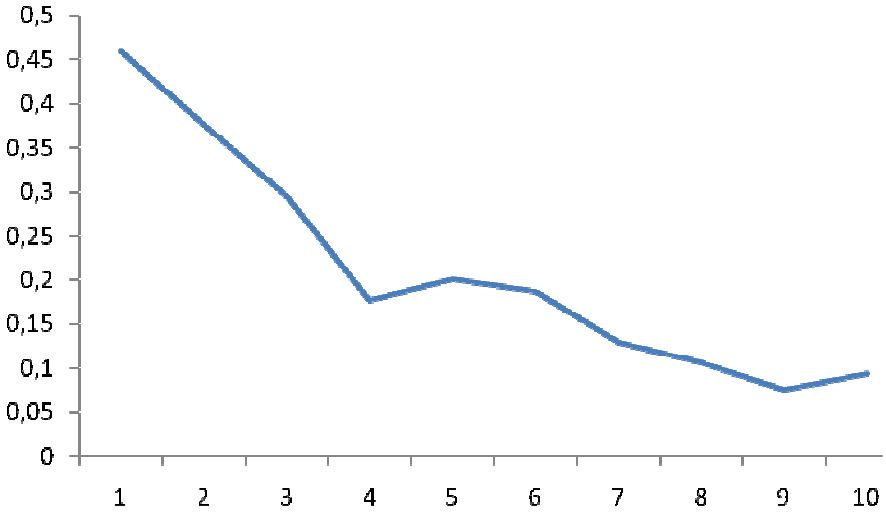


Figure 5, Fraction of cheaters by amount sent



Appendix: 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	€11.30	€13.85	€16.05	€17.90	€19.60	€21.20	€22.65	€24.05	€25.30

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:
  - A's earnings will be: € 10.50 – (euros sent to B) + (euros returned by B) – (€ 0.50 fee)
  - B's earnings will be: (euros received by B according to the table above) – (euros returned to A)

- If A sends nothing to B:
  - A's earnings will be € 10.50
  - B's earnings will be € 0.

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

1. Every participant specifies their decision for each possible role (A and B).
2. 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.
3. Within each pair, A's decisions will be combined with B's decision to determine the outcome for both A and B.

[Begin Experiment]

[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

[Sender decision screen 2]

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

__€ 1

__€ 2

…

__ € 10

[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 receiver € 21.20?

[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 : ____

…

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

[Belief elicitation instructions]

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.

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.

[Belief compensation formula]

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.

$$1 - \left(\frac{R - r}{r_{max}}\right)^2$$

- 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≤z, you will earn €5.00 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:

$$1 - \left(\frac{R-4}{10}\right)^2$$

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 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.
- On the other hand, the probability with which your estimate earns you €5 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

[Beliefs question 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: ___

[Beliefs question 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: ___

[Beliefs question 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: ___

[Beliefs question 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?: ___

[Beliefs question 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? ____