JENA ECONOMIC RESEARCH PAPERS

2010 – 090

Context and Interpretation in Laboratory Experiments: The Case of Reciprocity

by

Maria Vittoria Levati
Topi Miettinen
Birendra K. Rai

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

Jena Economic Research Papers 2010 - 090

Context and Interpretation in Laboratory Experiments: The Case of Reciprocity

Maria Vittoria Levati, Max Planck Institute of Economics, Jena, Germany

Topi Miettinen, Aalto University School of Economics, Helsinki, Finland

Birendra K. Rai[1], Monash Univeristy, Clayton, Australia

Abstract. The existing literature acknowledges that a mismatch between the experimenter's and the subjects' models of an experimental task can adversely affect the interpretation of data from laboratory experiments. We discuss why the two common experimental designs (between-subjects and within-subjects) used to conduct experiments may fail to sufficiently account for this concern. An alternative design for laboratory experiments is proposed which may alleviate this concern especially in studies of social preferences. The proposed design is used to answer some questions that have attracted continued attention in the literature on social preferences in general and reciprocity in particular.

Keywords: Experimental design; Context; Trust game

JEL codes: C70; C90; D63; D64

[1]Corresponding Author: birendra.rai@buseco.monash.edu.au

1. Introduction

One of the main reasons economists use laboratory experiments to answer questions of interest is that the experimenter can potentially exercise better control over some of the factors that affect the decisions of subjects. At the same time, experimenters are confronted with the possibility that the behavior of subjects is influenced by various contextual features which are often difficult to control in the lab. The analysis of data from a laboratory experiment not only involves documenting the responses of subjects but also providing an interpretation of the behavioral patterns. Orne (1962, 1973) highlighted that the behavior of subjects is a response to their perception of the experiment which may differ from what the experimenter believes it to be. As noted by Samuelson (2005, pp. 94, 96),

"... interpretation of experimental results can then depend importantly on how we imagine subjects perceive the game ... Despite an experimenter's best efforts to ensure that subjects understand what they are dealing with, ... it is not clear when we can be confident that the subjects' models match the experimenter's."

An important component of the requirement that a subject's model of a situation matches the experimenter's model is that the subject and the experimenter should attach the same *meanings* to the elements in the action sets of the agents involved in an interaction. A mismatch between the meanings attached by subjects and the experimenter to the various actions may adversely affect the interpretation of data, particularly in studies of social preferences. For example, Eckel and Wilson (2004) report that subjects within and across the roles of trustors and trustees differ in their perception of the trust game (Berg, Dickhaut, and McCabe, 1995).[2] If trustees construe the situation as that of receiving a short term loan from a friend, then they would not be inclined to return more than what was loaned to them by the trustors. Consequently, it may not be appropriate to interpret such a behavioral pattern as evidence for the lack of 'extra' returns to trust.

The question that motivates the present paper is: How to increase the likelihood that subjects attach the same meanings to actions as does the experimenter in any given lab experiment? Numerous researchers have alluded to the importance of coming up with experimental designs that can reduce the likelihood of such a mismatch and thereby aid in valid interpretation of lab data. Some of the prominent contributions include the work of Tversky and Kahnemann (1981) in the domain of individual choice, Dyer and Kagel (1996) in common-value auctions, Hoffman et al. (1996) in dictator games, and Levitt et al. (2010) in zero sum games. Ortmann and Gigerenzer (1997), Smith (2003, 2010), Camerer (2003), Samuelson (2005), Harrison and List (2004), Levitt and List (2007), Bardsley et al. (2010), and Bolton (2010) are excellent sources for a general discussion of this issue.[3]

[2] In the standard trust game the first mover (often referred to as the trustor) decides how much of his endowment to pass to the second mover. This amount is multiplied by a factor $m > 1$ by the experimenter, and the second mover (the trustee) then decides how much of this amount to return to the first mover. This game is believed to involve trust on part of the first mover as an own-payoff maximizing second mover would not return any amount and thus the first mover will be worse off by transferring any amount in the first place.

[3] The interested reader may also refer to a recent issue (Volume 73(1), 2010) of the Journal of Economic Behavior and Organization.

Consider an experiment that aims to provide evidence of a particular type of behavioral motivation by comparing the differences in behavior of subjects across two games. For instance, suppose a researcher hypothesizes that (i) the amount transferred by a dictator to the recipient in a dictator game[4] reflects her altruistic preferences and (ii) the amount transferred back by a trustee to the trustor in a trust game reflects the combined influence of altruistic and reciprocal preferences. Further suppose the main hypothesis of the researcher is that trustees will give more than dictators and this difference should be interpreted as evidence of a preference for reciprocity (Bolton et al., 1998 and Cox, 2004).

The mode of implementing the experiment is often decided after identifying the treatments that would allow for a test of the relevant hypotheses. At this stage, the experimenter has to make three decisions: which games subjects are *informed* about, which games they are allowed to *play*, and the *timing* of information and play.

The manner in which the games are implemented, along with the games themselves, determines the inferences one may draw from the data. One way to implement the above mentioned experiment would be to use a between-subjects design whereby one set of subjects would play the trust game and a different set of subjects would play the dictator game (and subjects who play one game will have no information about the other game). We shall argue that the between-subjects design may fail to sufficiently account for the concern regarding the interpretation of data in such experiments as there could be a significant mismatch between the experimenter's and subjects' models of *an* abstract game presented in isolation.

The motivation for using the within-subjects design, the leading alternative to the between-subjects design, comes from asking: What are the possible sources of differences in behavior *across* the two games in a between-subjects implementation? It prioritizes reducing differences in behavior across two games that could arise in the between-subjects implementation because of different subjects playing the two different games. Therefore, subjects are asked to play both the games in a sequential manner. However, this feature introduces some other concerns which have been acknowledged in the literature (Cox, 2010).

The motivation for the design proposed in the present paper comes from asking: What are the channels through which heterogeneities across individuals could lead to differences in behavior in the *same* game? Differences in preferences, differences in beliefs about others, and different perceptions of the same game are three obvious possibilities. The proposed design prioritizes reducing the differences in perception of the same game that could arise in the between-subjects implementation because of the abstract nature of the game. This design will be referred to as the *hybrid* design since it combines some features of between and within subjects designs. It involves

- providing information about both the games (the experimenter plans to compare in order to test the relevant hypotheses) to *all* subjects; but,
- randomly assigning some subjects to play *one* of the two games and the remaining subjects to play the other game.[5]

[4] In the basic version of the dictator game one subject is endowed by the experimenter with a certain amount of money and asked to share it with another subject (the recipient) who has no choice to make.

[5] Orne (1962, 1973) and Kahnemann (2003) hint towards such a design. However, no explicit discussion, or use, of such a design exists in the literature to the best of our knowledge.

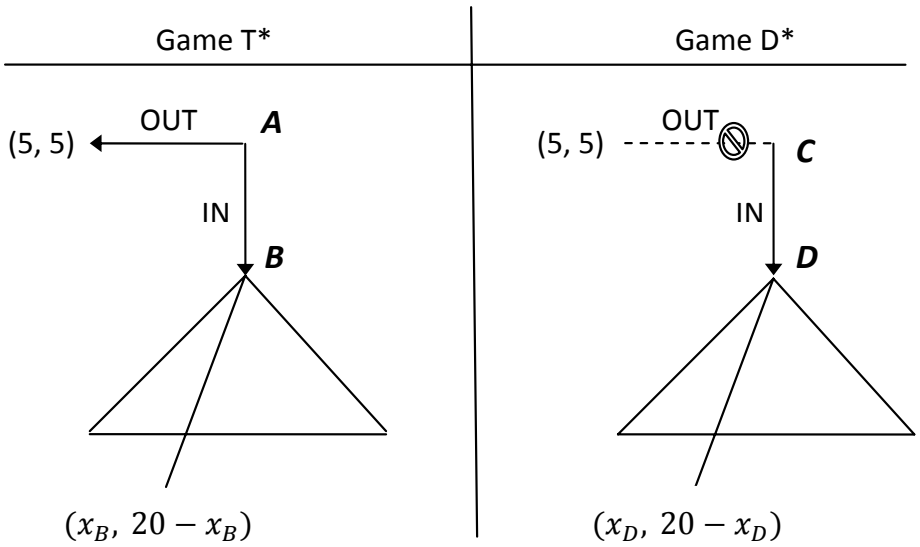

Figure 1: Trust and Dictator Games

The within-subjects design allows the experimenter to test whether the behavior in the second game is sensitive to having played the first game by conducting two different experimental sessions where (different) subjects play the two games in the two possible orders. Similarly, whether the behavior of subjects in one game is influenced by the information about the other game in a hybrid implementation can be easily tested.

Each mode of implementation accompanies certain tradeoffs. Section 2 provides a detailed discussion of these tradeoffs using the games illustrated in Figure 1. We choose these two games – modified versions of the trust and dictator games – as they are among the most widely studied games and help us clarify the tradeoffs associated with the three designs in a transparent manner. In the second part of the paper the choices and beliefs of subjects participating in a hybrid implementation of these games are analyzed to assess whether subjects perceive these games in the manner an experimenter would and to examine whether the data provide any evidence of reciprocal actions.

2. Tradeoffs in Experimental Designs

Consider the games T^* and D^* shown in Figure 1. In game T^*, subject A (the trustor) may choose IN or OUT. If A chooses OUT, then both subjects in the pair get a payoff of five. If A chooses IN, then B (the trustee) decides the amount x_B to give to A out of twenty. The game T^* possesses all the features of a trust game as (i) the choice of IN by the trustor leads to efficiency gains but (ii) an own-payoff maximizing trustee would not transfer any amount to the trustor if the trustor chooses IN.

The game D^* is essentially a dictator game presented in a slightly modified manner. Subject C (the recipient) has no freedom of choice as she must choose IN. Only subject D (the dictator) has an effective choice in this game and has to decide the amount x_D to give to subject C out of twenty. The games T^* and D^* may also be thought of as voluntary and

involuntary trust games, respectively (McCabe et al., 2003; Smith, 2003). The action set faced by a trustee in game T^* is identical to the action set faced by a dictator in game D^*.

The experimenter designs the two games such that the only difference between them is whether the first mover has a choice or not. Given the salience of this difference, she may find it reasonable to interpret (i) the voluntary choice of IN by a trustor as truly a move that involves trust, (ii) the choice faced by a trustee as involving reciprocation, and to be confident that (iii) the choice faced by a dictator differs from the choice faced by a trustee because it lacks any element of reciprocation.

The question regarding which mode of implementation should be used to test whether there is any evidence of reciprocal actions/preferences arises *after* the experimenter has settled on the appropriate treatments. We now discuss the tradeoffs involved in a between-subjects, within-subjects, and hybrid implementation of games T^* and D^*.

2.1. Between-subjects design

A between-subjects implementation of the games T^* and D^* illustrated in Figure 1 would involve one set of subjects playing the game T^* without any information about the game D^* and a different set of subjects playing D^* without any information about T^*.[6]

Let X_B and X_D denote the vectors of amounts given by trustees to trustors and by dictators to recipients, respectively. One may *define* transfers by dictators as a measure of their altruistic preferences. Similarly, one may define transfers by trustees as reflecting the combined influence of their altruistic and reciprocal motivations. Then, any observed positive difference between the transfers by trustees and dictators will, by definition, capture positive reciprocity. But, can we unambiguously *interpret* the observed difference between X_B and X_D as evidence of positive reciprocity.

Cherry et al. (2002), Dana et al. (2005), List (2007), and Bardsley (2008) provide compelling evidence that interpreting the transfers by dictators as evidence of context-independent altruistic preferences may not be appropriate. Eckel and Wilson (2004) would caution against interpreting the behavior of trustees as involving reciprocation unless we are confident that subjects indeed view the game T^* as involving trust and reciprocity. If it is difficult to interpret the meaning of X_D and X_B, then interpreting the difference between X_B and X_D as evidence of reciprocal preferences seems suspect.

It may be helpful to ask what makes the researcher confident regarding the plausibility of the hypothesis that the difference in the behavior of trustees and dictators can be interpreted as evidence of positive reciprocity. The answer lies in recognizing that the researcher compares the two games while formulating the hypothesis. Thinking about both the games *simultaneously* provides the researcher the appropriate context to attach a largely unambiguous meaning to the elements in the action sets of the agents playing the games.[7]

[6]The between-subjects design essentially mimics the experimental procedure used in physical sciences where the aim is to document differences in responses to different stimuli. The main difference between experiments in physical sciences and economics is that the experimenter tries to ensure that subjects in an economic experiment understand the task they have been presented.

[7]Birnbaum (1999) shows that it may be unreasonable to assume that the context is the same across the different sets of subjects participating in the between subjects implementation.

Subjects playing the trust game are not informed about the dictator game, and vice-versa, when one uses a between-subjects design whereby each game is played out in different experimental sessions by different sets of subjects. There is, therefore, an asymmetry between the context in which the researcher formulates the hypothesis and the context in which subjects participating in a between-subjects implementation make their choices.

This asymmetry will not create any problems if subjects attach the same meanings to the various actions available to the subjects in different roles in the trust/dictator game presented to them in isolation. The very process of understanding the rules of *one* abstract game might (consciously or unconsciously) lead subjects to form analogies between the task they have been presented with and situations they have encountered outside the laboratory (Smith, 2003). Thus, in the between-subjects implementation, it becomes difficult to dismiss the methodological concern which has been raised by numerous researchers and summarized by Harrison and List (2004, pp. 1050) as follows.

... experimenters should be wary of the conventional wisdom that abstract, imposed treatments allow general inferences. In an attempt to ensure generality and control by gutting all instructions and procedures of field referents, the traditional lab experimenter has arguably lost control to the extent that subjects seek to provide their own field referents.

Moreover, when subjects seek to provide their own field referents different subjects may construct different models of the same situation which may severely limit our ability to provide a coherent interpretation of the behavior of subjects. This critique questions the interpretability of data obtained via lab experiments. However, it may be worth noting that it does not regard abstract experimental treatments and the lack of field referents as the fundamental problem. A constructive reading of this critique essentially points out the importance of designs for lab experiments which ensure better task perception by subjects and reduce the need for subjects to rely on personal field referents and thereby decrease the likelihood of a mismatch between the experimenter's and the subjects' models of the interaction.

Framing the experimental task using a particular context can reduce this mismatch in various cases but it comes at the cost of reduced generality of results (Ortmann and Gigerenzer, 1997). We shall argue that the hybrid design can potentially reduce the mismatch without compromising much with generality.

2.2. Within-subjects design

The standard within-subjects implementation of the games illustrated in Figure 1 would involve subjects receiving instructions about one of the two games (say, T^*), playing that game, receiving the instructions for the second game, and then playing the second game. Those who play as trustees (trustors) in T^* would play as dictators (recipients) in D^*. When subjects would be provided the detailed instructions for the first game they would be informed that they will have to play another game which will be described in detail after the first game has been played. Subjects would be paid for only one of the two games to reduce income effects and to increase the likelihood that they treat each game as a separate game.

The prime motivation for using a within-subjects design, instead of a between-subjects design, is to control for individual heterogeneities. In addition, it gathers more data per subject. The appeal of within-subjects implementation for any given experiment ultimately depends upon how the experimenter subjectively resolves the tradeoff between these two major benefits and the concerns discussed below.

The informational environment faced by subjects while making their decisions in the *first* game in the within-subjects implementation differs in a crucial way from that faced by subjects in a between-subjects implementation. The mere fact that subjects know that they will be playing another game may influence their behavior (Smith, 2010). It may adversely affect the interpretability of data by creating a mismatch between the experimenter's and subjects' models of the situation.

Subjects are usually paid for only one of the several games they play during a within-subjects implementation to ensure that subjects treat each game as a separate game. However, it is difficult to verify that subjects do treat each game as a separate game.[8] Cox (2010) provides a detailed discussion of the payment mechanism used in within-subjects designs by explaining its theoretical underpinnings and highlighting the associated concerns.

While using a within-subjects implementation one usually tests for order effects: whether or not the behavior of subjects in either game is sensitive to having played the other game. If there is evidence of order effects, then the experimenter may have to abandon the data and carry out the experiment using a between-subjects design (Bohnet et al., 2008). It is not clear what inference can be drawn in such a scenario.

The discussion of the between-subjects design in the previous section suggests that subjects may better understand the meanings of actions available in the two games only after they are informed about the second game during a within-subjects implementation. This could be one of the reasons for order effects.

2.3. Hybrid design

A hybrid implementation of the games T^* and D^* illustrated in Figure 1 would be as follows. All subjects in a given session would be informed about *both* the games. Subjects would then be randomly assigned to one of the four player-roles. A subject would get paid only for the game she actually plays. It will be useful for the following discussion to take note of a *variant* of the hybrid design wherein subjects would be informed about both the games but only one of the two games would be selected for actual play during an experimental session.

The main difference between the hybrid design and its variant on the one hand, and the existing designs on the other, pertains to the information subjects have while making decisions. This is obviously the case in comparison with a between-subjects implementation. Even if subjects in a within subjects implementation were to be informed about all their tasks at the very beginning, it would still not be similar to the hybrid design because each subject would participate in only one game in the hybrid implementation.

[8] Blanco et al. (Forthcoming) point towards this possibility as one of the likely reasons behind the finding that the theory of inequity-aversion (Fehr and Schmidt, 1999) receives support at the aggregate level but not at the individual level.

Our aim will be to clarify the implicit and the explicit tradeoffs in the hybrid design using the games T^* and D^*. We divide the discussion into three parts: issues related to the perception of the games, the potential for experimenter demand effects, and the logistical (dis)advantages. In general, we shall argue that any feature of the hybrid design which may be regarded as a 'conceptual drawback' is also present in the between-subjects or/and the within-subjects design.[9]

2.3.1. Perception

Wilson (2010, pp. 81) asserts that

"... if we as the experimenters cannot peer into the minds of our subjects to ascertain their social motives, then our subjects face the same uncertainty when interacting in the tasks that we give them. Thus, the problem boils down to one of agreement: Do the subjects agree on the context of the interaction: Do the subjects agree that this situation invokes the social motives of a rule of reciprocity, ..., or some other rule to guide behavior?"

By providing the same informational context as used by the researcher the hybrid design could help in reducing the likelihood of disagreement among the subjects, and between the experimenter and the subjects, regarding the meanings of the actions available to subjects. Consequently, one would expect the likelihood of a mismatch between the experimenter's and subjects' models of the situation to be lower.

If the experimenter believes that the games T^* and D^* are appropriate for identifying reciprocity, then it will be inconsistent to simultaneously believe that the mismatch between the experimenter's and subjects' models of the interactions will be greater in the hybrid design compared to the between-subjects design. After all, the choice of the games reflects the conviction of the experimenter that they are appropriate for examining the hypotheses.

It is possible that there will be a greater cognitive load on subjects in the hybrid implementation and the within-subjects implementation compared to a between-subjects implementation. However, greater cognitive load in itself may not be a cause for concern if it ultimately allows the subjects to better understand the experimental task.[10]

2.3.2. Demand effects

The critical issue facing the hybrid design, or any other design, that reduces the 'mismatch' is that it will be subject to the critique of *experimenter demand effects*. However, we shall

[9] In a within subjects implementation, the choice in one game may influence the choice in a latter game (*anchoring*) and thereby lead to order effects in the decisions of subjects. In the hybrid design, information about the other game may influence the decision of a subject in the game she is actually playing. The current practice in experimental economics seems to be that the results from a within-subjects implementation are considered useful only if there is no significant evidence of anchoring in subjects' decisions. Whenever there is no such concern for the within-subjects implementation where subjects get to know *and* play both games, there should be little cause for concern in the hybrid implementation where subjects would get to know both games *but* will play only one of the two games.

[10] We thank Andreas Ortmann for clarifying the issue pertaining to cognitive load.

argue that this concern does not hold up under close scrutiny. Also note that the within-subjects implementation is also subject to a similar critique arising from subjects playing the games in a sequential manner. The standard counter-argument in the case of within-subjects implementation is that one can test for the effects of playing the games sequentially by conducting a different experimental session where the games are played in the reverse order. As discussed below, the hybrid design also allows testing for the effects of information about the other game in a hybrid implementation.

Suppose we find a significant difference in the behavior of trustees in T^* and dictators in D^* using a hybrid implementation. A possible way to question the validity of such a finding would be as follows.

1. *Subjects in the role of dictators in D^* may believe that it is the trustees who are supposed to give relatively more in this experiment.*

2. *Therefore, the dictators might give less than what they would have given if they did not have any information about the game T^*.*

Statement [2] implicitly presumes that there is something akin to the 'true' dictator behavior which would be observed when dictators make their decisions without any information about the trust game. Apart from its obvious logistical advantage, it does not provide any conceptual reason to treat the behavior in the isolated dictator game as the appropriate benchmark.

The critique of demand effects implicitly favors the between-subjects design as the correct mode of implementing an experiment without a sound conceptual justification. As argued by Zizzo (2010), demand effects can not be ruled out in the between-subjects implementation. In fact, the between-subjects design often provides no clue regarding the source, nature, or the impact of demand effects. It also distracts from the key point that we may not be interested in the behavior in a game, *per se*, as the game primarily serves as the tool that allows us to explore an underlying question. Unless the research question is understanding the behavior in a particular game, the between-subjects design has no special conceptual significance.

Nonetheless, whether the information about the other game influences the behavior of subjects in the game they are playing can be easily tested. Consider statement [2]. There would be two possibilities to consider. First, merely the information about the trust game might influence the behavior of dictators. Second, there might be an additional impact due to some subjects actually playing the trust game in the same experimental session. The presence of some subjects playing as trustees might increase the likelihood of dictators thinking as in [1] compared to the case where subjects are informed about both the games but all subjects are asked to play the dictator game.

The first possibility can be tested by comparing the behavior of dictators in a between-subjects implementation with that of dictators in the *variant* of the hybrid design described earlier. The second possibility can be tested by comparing the behavior of dictators in the hybrid implementation and its variant.

Demand effects create interpretational problems if subjects can come to a specific conclusion regarding the actions that would be in line with the expected objectives of the experimenter and *act* as if to please the experimenter (Zizzo, 2010). The extent to which subjects can

come to a conclusion regarding the actions that would support the experimenter's hypotheses in the hybrid implementation will be no more than in the within-subjects implementation after subjects receive information about the second game. Moreover, there seems to be little a priori reason to believe that subjects can come to such a conclusion in *all* lab experiments (Lambdin and Shaffer, 2009).[11]

The concern regarding demand effects may be phrased in two other ways. First, it may be argued that the hybrid design may reveal the research question to the subjects. However, if the research question gets revealed in the hybrid implementation of an experiment, then it will eventually get revealed in the within-subjects implementation as well. Even if a design does reveal the question, it is difficult to accept that revealing 'the' question is unambiguously worse than revealing 'some' question given that economic experiments attempt to examine behavior in the presence of well-defined incentives.

Second, subjects' perceptions may be affected by the possibility that they might believe that the experimenter sees a difference between the two games being used in the hybrid implementation even if the subjects themselves do not see any difference. Once again, there is hardly anything additional in the within-subjects implementation relative to the hybrid implementation to stop the subjects from thinking in this manner. Thus, the fact that subjects have information about both the games in the hybrid implementation need not be as troublesome as it appears.

2.3.3. Logistics

On the positive side, the hybrid design exploits the full force of randomization as subjects are randomly assigned to the various player-roles during each experimental session which may help reduce practical concerns regarding different types of subjects showing up for different experimental sessions.

Each subject receives exactly the same instructions in the hybrid implementation which is not the case in a between-subjects implementation. The hybrid implementation (but not its variant) allows us to credibly elicit incentivized beliefs of a subject about the game she is playing herself (the within-game beliefs) and about the other game being played by some of the other subjects (the cross-game beliefs) during the same experimental session.

The hybrid implementation would consume at least as much time per experimental session and would provide no improvement in terms of the data collected per session compared to the between or within subjects designs. It would be difficult to implement if the overall experiment involves a large number of treatments. However, as hypothesis tests are often conducted by comparing behavior across two treatments, pairs of treatments that allow a particular hypothesis to be tested could be implemented in a session.

[11] Ananish Chaudhuri has provided us the results from an ongoing study wherein subjects are provided a clear cut 'explanation' of the strategic structure of a trust game. One might argue that this alteration introduces demand effects. However, it is not clear whether this will lead to greater or lower trust and trustworthiness on part of the trustors and trustees, respectively. Using a between-subjects design he finds greater trust and trustworthiness in the 'explanation' treatment relative to the baseline treatment that was otherwise identical to the treatment with the explanation.

3. An application of the hybrid design

We next report the results from the hybrid implementation of the games T^* and D^* and compare the results with the existing literature. Our interest lies in examining the choices and beliefs of subjects in the hybrid implementation of games T^* and D^* to understand, among other things, whether (and, which) subjects perceive the choice faced by a trustee to be different from the choice faced by a dictator by subjects who actually play only the trust game or only the dictator game.

3.1. Participants and procedure

We conducted five sessions each consisting of 32 subjects in the experimental laboratory at the Max Planck Institute of Economics, Jena (Germany). Subjects were students at the University of Jena. In each session, subjects first read the instructions on their own and then the instructions were read aloud by an experimenter. The instructions were framed in a neutral language and explained *both* the games to all the subjects. A comprehension quiz containing questions related to both games was administered to ensure subjects understood both the games. Subjects were then randomly assigned to one of the four possible roles in the experiment: trustors and trustees in T^* and recipients and dictators in D^*.

The games were played only once. Each session contained eight pairs of subjects playing T^* and another eight pairs of subjects playing D^*. The data consists of observations on a total of 160 subjects – 40 subjects in each of the four roles. Each subject in the role of a trustor in T^* had to choose IN or OUT and each subject in the role of a trustee had to decide how much to give to the trustor out of twenty monetary units. Every trustee made the choice without knowing the choice made by the trustor. Casari and Cason (2009) show that trustees give less to trustors when they do not know the choice made by the trustor. This should make it difficult for us to find differences in the choices of trustees and dictators.

The within and cross game beliefs of subjects were elicited (without prior announcement) after they made their choices. The belief elicitation was incentivized to limit the possibility of ex-post rationalization. Note that this concern should be lower for cross-game beliefs. Subjects were paid three monetary units for an accurate guess, one if their guess deviated (in absolute value) by one unit, and nothing otherwise. We first elicited the within-game beliefs and then the cross-game beliefs for subjects in each role. Table 1 lists the exact sequence of choices and beliefs for subjects in each role.

Subjects were informed about their payoffs only after they had made their choices and all the beliefs had been elicited. Each subject was paid the earnings from the experiment and a show-up fee of 2.50 monetary units in cash. Each session lasted nearly an hour (including the instructions and payment) and the average earnings were about fourteen monetary units.

The notation we shall use throughout the paper is as follows. The vector of choices by subjects in role i will be denoted by X_i. The vector of (first order) beliefs of subjects in role i about the behavior of subjects in role j shall be denoted by F_{ij}. The vector of (second order) beliefs of subjects in role i regarding the belief of a subject in role j about the behavior of a subject in role k shall be denoted by S_{ijk}. The analysis of data follows.

Table 1: Sequence of Choices and Beliefs

Notation	**Description**
A: Trustors	
$x_A \in \{IN, OUT\}$	Choose In or OUT.
$f_{AB} \in \{1, ..., 19\}$	Guess how much the B in your pair gave to you.
$s_{ABA} \in \{0, ..., 8\}$	Guess the number of A's that chose IN according to the B in your pair.
$f_{AA} \in \{0, ..., 7\}$	Guess how many of the 7 A's other than you chose IN.
$f_{AD} \in \{1, ..., 19\}$	Guess the amount given by a randomly chosen D.
B: Trustees	
$x_B \in \{1, ..., 19\}$	How much you want to give to the A in your pair?
$f_{BA} \in \{0, ..., 8\}$	Guess the number of A's that chose IN.
$s_{BAB} \in \{1, ..., 19\}$	Guess the amount the A in your pair thinks you gave him.
$f_{BB} \in \{1, ..., 19\}$	Guess the amount given by a randomly chosen B other than you.
$f_{BD} \in \{1, ..., 19\}$	Guess the amount given by a randomly chosen D.
C: Recipients	
x_C	You do not have to make any choice.
$f_{CD} \in \{1, ..., 19\}$	Guess how much the D in your pair gave to you.
$f_{CA} \in \{0, ..., 8\}$	Guess the number of A's that chose IN.
$f_{CB} \in \{1, ..., 19\}$	Guess the amount given by a randomly chosen B.
$s_{CAB} \in \{1, ..., 19\}$	Guess how much a randomly chosen A thinks the B in his pair gave to him.
D: Dictators	
$x_D \in \{1, ..., 19\}$	How much you want to give to the C in your pair?
$s_{DCD} \in \{1, ..., 19\}$	Guess the amount the C in your pair thinks you gave him.
$f_{DD} \in \{1, ..., 19\}$	Guess the amount given by a randomly chosen D other than you.
$f_{DA} \in \{0, ..., 8\}$	Guess the number of A's that chose IN.
$f_{DB} \in \{1, ..., 19\}$	Guess the amount given by a randomly chosen B.

3.3. Choices

Nineteen out of forty trustors (roughly 50%) chose to trust the trustee in game T^* by choosing IN. As depicted in Table 2, fifteen of these trustors expected to receive strictly more than the outside option of five, two expected to get exactly five, and the remaining two expected to receive strictly less than five. Thus, roughly 90% of the trustors who chose IN expected to receive an amount (weakly) greater than their outside option. This finding contrasts starkly with the results of Ashraf et al. (2006) who find that only 62% of trustors in an investment game expect to get back (weakly) more from the trustee than what they transferred.

Out of the 21 trustors who chose OUT, twelve expected to receive strictly less than five, five expected to receive exactly five, and four expected to receive strictly more than five. Thus, roughly 80% of the trustors who chose OUT expected to receive an amount (weakly) lower than their outside option.

Table 2: Choices of Trustors in T^*

Choice	#	# ($f_{iB} < 5$)	# ($f_{iB} = 5$)	# ($f_{iB} > 5$)
IN	19	2	2	15
OUT	21	12	5	4

The distribution of choices of trustees in game T^* and dictators in game D^* are summarized in Table 3 and illustrated in Figure 2. Exactly 50% of the trustees (20 out of 40) gave strictly more than the outside option of five to trustors. The median and mean amounts given by trustees are greater than the corresponding values for dictators.

Table 3: Choices of Trustees in T^* and Dictators in D^*

	Min	1^{st} Qu.	Median	Mean	3^{rd} Qu.	Max.	St. Dev.
X_B	1.00	2.00	5.50	5.90	9.00	18.00	4.09
X_D	1.00	1.00	2.00	3.55	5.00	15.00	3.42

Our first hypothesis tests for differences in the choices of trustees and dictators.

H1: $X_B = X_D$

Statistical tests reveal a significant difference in the amounts given by trustees and dictators. The p-values corresponding to Kolmogorov-Smirnov test and Wilcoxon Rank Sum test are 0.029 and 0.007, respectively.

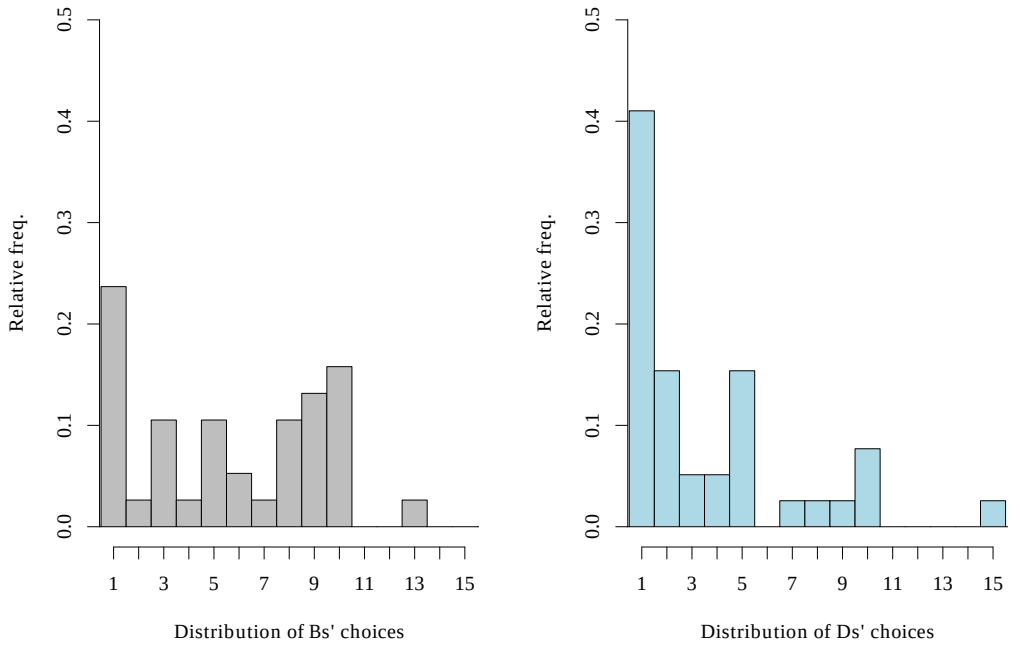

Figure 2: Choices of Trustees and Dictators

This result suggests that subjects respond more generously when they know that their action set will be realized from a voluntary choice, rather than an involuntary choice, of another subject. Though the result is not surprising, it differs from the results obtained by Dufwenberg and Gneezy (2000) in an almost identical game.[12]

There are two distinct pieces of information in this result. First, the behavior of dictators is *different* from that of trustees. Second, trustees give *more* than dictators on average. Bolton et al. (1998) and Cox (2004) provide one plausible interpretation of this finding as described earlier. Following Wilson (2008, 2010) one could also suggest that trustees use the 'rule of reciprocity' and dictators use, say, the 'rule of desert' with the former being stronger than the latter. Our experiment can not distinguish between these two explanations.

McCabe et al. (2003) find a significant difference (using a between-subjects design) in the choices of trustees and dictators when a trustee decides after observing the choice of the trustor. They conjecture that Dufwenberg and Gneezy (2000) do not find any significant difference in the behavior of trustees and dictators probably because each trustee makes her decision without knowing the decision of the trustor. The results from the hybrid design do not seem to support this conjecture. We believe that having information about both the games allows subjects to associate a meaning to their choices in a way that is not possible in the between-subjects implementation. This in turn leads to behaviors and beliefs that accord with our intuitive understanding of reciprocity.

[12]The payoff to the trustee in the trust game used by Dufwenberg and Gneezy (2000) resulting from the OUT choice by the trustor is zero. We set the corresponding payoff at 5 euros. If anything, this should make a trustee in T^* feel less gratitude towards a trustor who chooses IN compared to a trustee in their trust game.

3.4. Beliefs about choices of subjects in a given role

The summary statistics regarding the beliefs of subjects in various roles are tabulated in Appendix I. In this section we wish to examine whether assignment to a particular role influences subjects' beliefs. In order to do so we compare the beliefs of subjects in any two roles about the choices made by subjects in a particular role.

H2: $F_{ik} = F_{jk},\ i,j \in \{A,B,C,D\},\ i \neq j,\ k \in \{A,B,D\}$

Tables 4, 5, and 6 report the p-values of Kolmogorov-Smirnov (KS) and Wilcoxon Rank Sum (WRS) tests. The p-values are greater than 0.1 in 35 out of the 36 cases as reported in these tables. The only significant difference is found in a comparison of the beliefs of trustees and dictators regarding the choices of trustors using the WRS test (p-value = 0.04). We shall see in the following sections that these findings hide the systematic differences in beliefs across subjects in the same role who behave differently.

Table 4: Beliefs about choices of trustors

p-value† of	F_{AA} vs F_{BA}	F_{AA} vs F_{CA}	F_{AA} vs F_{DA}	F_{BA} vs F_{CA}	F_{BA} vs F_{DA}	F_{CA} vs F_{DA}
KS	0.401	0.573	0.573	0.759	0.263	1.000
WRS	0.284	0.636	0.395	0.108	**0.040**	0.630

† Each trustor was asked to guess how many out of the 7 trustors other than herself (in the experimental session) chose IN. Subjects in all the other roles were asked to guess how many out of the 8 trustors chose IN. To carry out the statistical tests the reported beliefs have been 'normalized' to lie between 0 and 1.

Table 5: Beliefs about choices of trustees

p-value of	F_{AB} vs F_{BB}	F_{AB} vs F_{CB}	F_{AB} vs F_{DB}	F_{BB} vs F_{CB}	F_{BB} vs F_{DB}	F_{CB} vs F_{DB}
KS	1.000	1.000	0.914	0.988	0.759	0.914
WRS	0.977	0.961	0.508	0.992	0.554	0.524

Table 6: Beliefs about choices of dictators

p-value of	F_{AD} vs F_{BD}	F_{AD} vs F_{CD}	F_{AD} vs F_{DD}	F_{BD} vs F_{CD}	F_{BD} vs F_{DD}	F_{CD} vs F_{DD}
KS	0.988	0.988	0.401	0.988	0.914	0.988
WRS	0.946	0.632	0.390	0.658	0.559	0.780

3.5. Beliefs of trustors and trustees conditioned on choice

The choice of IN by a trustor in T^* is qualitatively different from the choice of OUT. Similarly, the choice of giving more than the outside option by a trustee in T^* may be regarded as qualitatively different from the choice of giving no more than the outside option. Let us denote the trustors who chose IN as A_I's and the trustors who chose OUT as A_o's; and, the trustees who gave strictly more than the outside option as B_H's and those who gave no more than the outside option as B_L's.

We now investigate whether there is any statistical difference in the within and cross game beliefs of the two sets of trustors and the two sets of trustees.

H3.1: $F_{A_Ik} = F_{A_ok},\ k \in \{A, B, D\}$ and $S_{A_IBA} = S_{A_oBA}$
H3.2: $F_{B_Hk} = F_{B_Lk},\ k \in \{A, B, D\}$ and $S_{B_HAB} = S_{B_LAB}$

Tables 7 and 8 list the p-values corresponding to the tests of H3. In general, trustors who chose IN may be termed *optimistic* relative to the trustors who chose OUT. Compared to trustors who chose OUT, the trustors who chose IN (i) expect trustees to give relatively more, (ii) believe that trustees expect a relatively greater number of trustors to choose IN, (iii) expect a relatively greater number of other trustors will choose IN, and (iv) expect a randomly chosen dictator to give relatively more. The differences in beliefs turn out to be weakly significant using at least one of the two tests. Similarly, trustees who gave more than the outside option may be termed optimistic relative to the trustees who gave no more than the outside option. Figures 3 and 4 illustrate the various beliefs of trustors and trustees conditioned on their choices.

Table 7: Comparing beliefs of A_I's and A_o's

p-value of	F_{A_IB} vs F_{A_oB}	S_{A_IBA} vs S_{A_oBA}	F_{A_IA} vs F_{A_oA}	F_{A_ID} vs F_{A_oD}
KS	**0.002**	*0.091*	**0.005**	0.611
WRS	**<0.001**	**0.019**	**<0.001**	*0.088*

Table 8: Comparing beliefs of B_H's and B_L's

$p-$value of	F_{B_HA} vs F_{B_LA}	S_{B_HAB} vs S_{B_LAB}	F_{B_HB} vs F_{B_LB}	F_{B_HD} vs F_{B_LD}
KS	0.329	**0.005**	**<0.001**	**0.001**
WRS	*0.073*	**0.010**	**<0.001**	**<0.001**

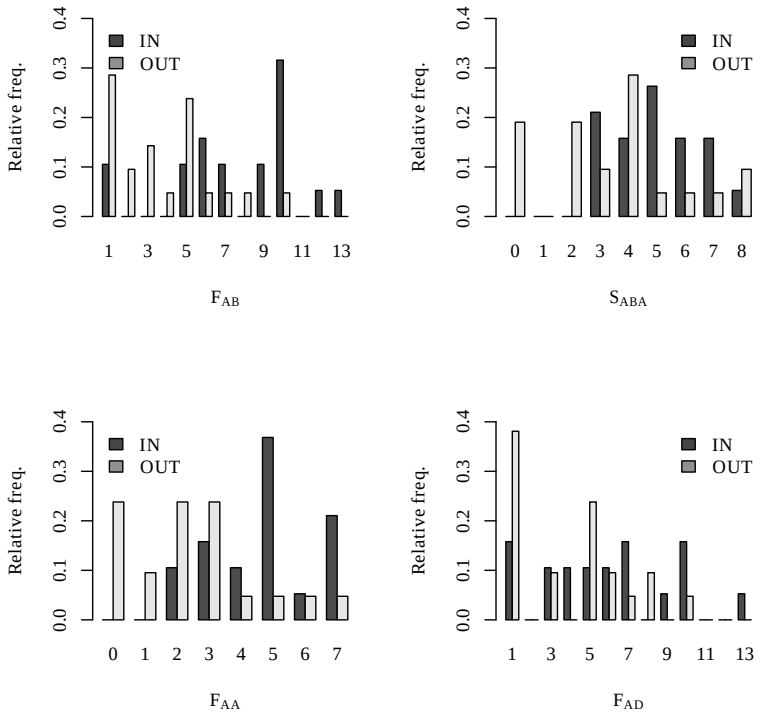

Figure 3: Beliefs of trustors conditioned on choice

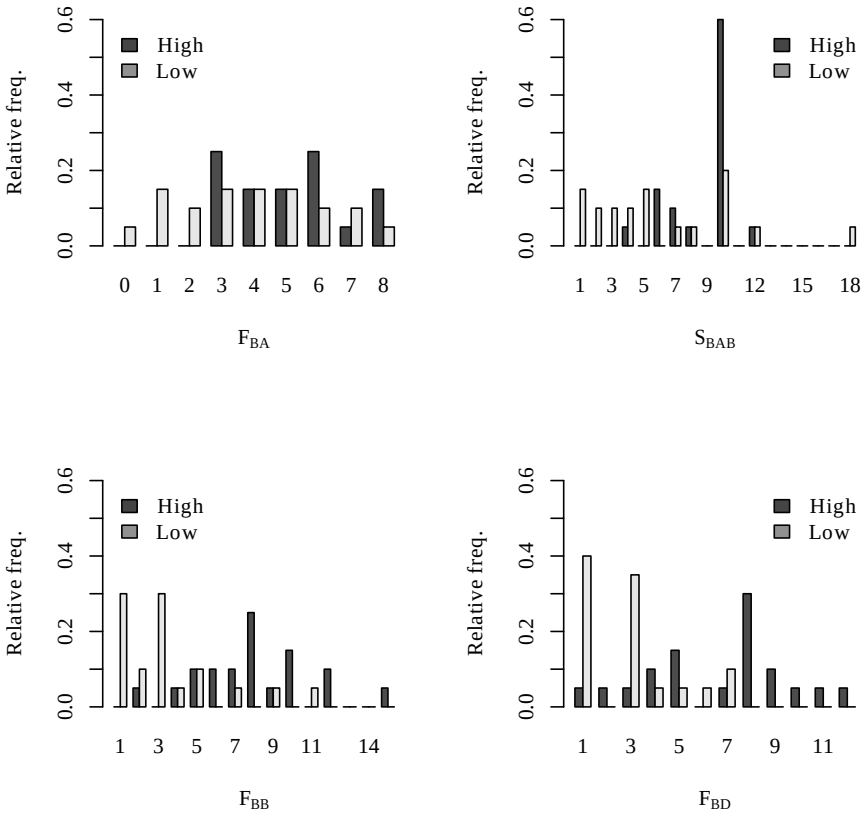

Figure 4: Beliefs of trustees conditioned on choice

The dominant approach to explain pro-social actions is to model agents as having (conditional or unconditional) other regarding preferences. Heterogeneities across agents are captured by allowing for differences in one or more preference parameters and assuming that the distribution of these parameters is common knowledge among the agents (for example, Levine 1998; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Sobel, 2005). Other researchers additionally allow for differences in prior beliefs in the specification of types of agents (Ellingsen and Johannesson, 2008). Systematic differences in the within and cross game beliefs of trustors and trustees who behave differently provide support for the latter modeling strategy irrespective of whether or not such a pattern of choices and beliefs is labeled as a manifestation of the false-consensus effect (Engelmann and Strobel, 2000).

3.6. Beliefs about the choices of dictators and trustees

It has already been reported that the choices of trustees are significantly different from the choices of dictators. We now examine whether subjects expect a difference in the choices of trustees and dictators. Table 9 provides a summary of the beliefs of subjects in each role regarding the behavior of trustees and dictators. For example, 8 out of the 40 trustors expect a randomly chosen dictator to give strictly more than the trustee they are paired with, 17 trustors expect both to give equal amounts, and the remaining 15 trustors expect the trustee to give more. The average of the difference between the individual level beliefs of trustors regarding the choices of trustees and dictators is 0.75. Apart from the trustors who chose out (A_o's) the number of subjects in any given role who expect trustees to give relatively more is greater than the number of subjects who expect dictators to give relatively more. Figure 5 illustrates the beliefs of subjects across all the roles regarding the choice made by a randomly chosen trustee and a randomly chosen dictator.

Table 9: Beliefs about choices of trustees and dictators

	A	A_I	A_o	B	B_H	B_L	C	D	Total
Observations	40	19	21	40	20	20	40	40	160
$\#(f_{iB} < f_{iD})$	8	2	6	9	4	5	6	9	32
$\#(f_{iB} = f_{iD})$	17	6	11	13	5	8	15	13	58
$\#(f_{iB} > f_{iD})$	15	11	4	18	11	7	19	18	70
$\overline{f_{iB} - f_{iD}}$	0.75	1.84	-0.24	0.90	1.25	0.55	0.925	0.65	0.81

We first pool the data on subjects in *all* the four roles and test the following hypothesis using the (paired) Wilcoxon Signed Rank test.

H5: $F_{allB} = F_{allD}$

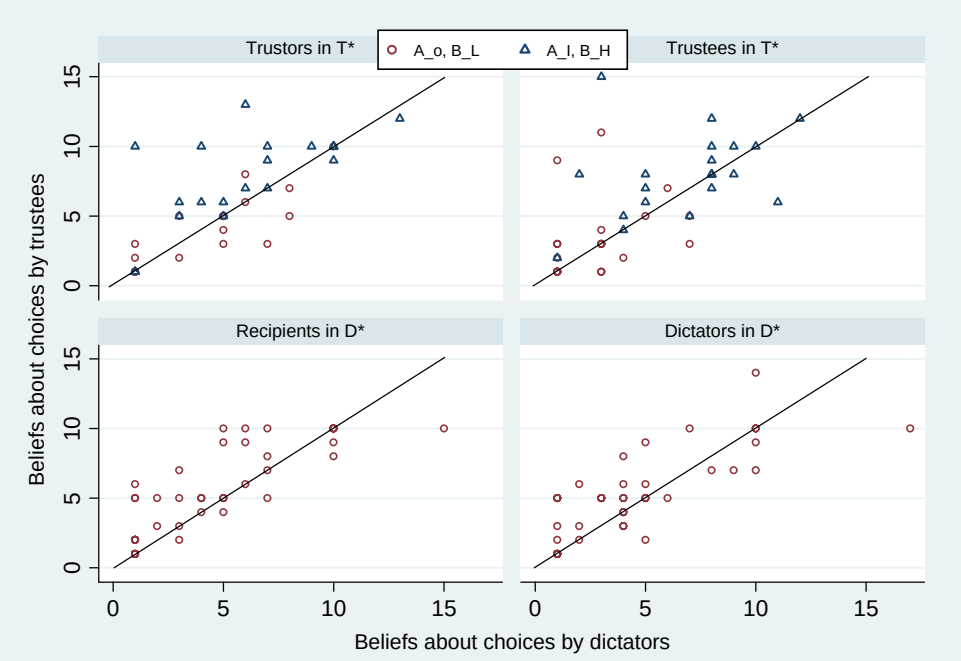

Figure 5: Beliefs about choices of trustees and dictators

The p-value corresponding to the Wilcoxon Signed Rank test is less than 0.001 and thus the above hypothesis is strongly rejected. We next examine whether this result is primarily driven by subjects in a particular role or do subjects in each role expect a difference in the choices of trustees and dictators.

H6: $F_{kB} = F_{kD},\ k \in \{A, A_o, A_I, B, B_L, B_H, C, D\}$

The p-values corresponding to Wilcoxon Signed Rank tests under the above hypotheses are reported in Table 10. The null hypothesis is strongly rejected for recipients and dictators. It is also weakly rejected for trustors. After dividing the trustors according to their choice we find that A_I's expect a significant difference in the choices of trustees and dictators whereas A_o's do not. A roughly similar qualitative pattern is observed in the case of trustees. Although we fail to reject the null hypothesis for trustees as a whole, this result seems to be primarily driven by those trustees who gave no more than the outside option (B_L's). They do not expect a significant difference in the choices of trustees and dictators but B_H's do expect a (weakly) significant difference.

Table 10: Tests of H6

p-value of	F_{AB} vs F_{AD}	F_{A_oB} vs F_{A_oD}	F_{A_IB} vs F_{A_ID}	F_{BB} vs F_{BD}	F_{B_LB} vs F_{B_LD}	F_{B_HB} vs F_{B_HD}	F_{CB} vs F_{CD}	F_{DB} vs F_{DD}
WSR	*0.059*	0.606	**0.006**	0.159	0.842	*0.089*	**0.011**	**0.040**

A subject may believe that the choice faced by a trustee differs from the choice faced by a dictator because the former involves reciprocation but the latter does not. However, such a subject may not be confident that other subjects will share this view or that even if subjects share this view they may not act in accordance with this view. Among subjects in the roles of trustors and trustees it is only those who behave in a pro-social manner (A_I's and B_H's) that expect a difference in the choices of trustees and dictators. It might be reasonable to conclude that A_I's and B_H's, the recipients, and the dictators

1. perceive the nature of the choice faced by trustees to be different from that faced by dictators;

2. believe that other subjects in the roles of trustees and dictators will also perceive this difference; and,

3. are confident that the choices of trustees and dictators will reflect this difference.

The lack of difference in the beliefs of A_o's and B_L's regarding the choices of trustees and dictators reflects, at the least, the failure of (3). Given that the remaining subjects seem to perceive a difference in the choices faced by trustees and dictators we suspect that this does not imply a failure of (1) for A_o's and B_L's. It may not be unreasonable to suggest that subjects across all the roles perceive a difference between the choice faced by the trustees and the choice faced by the dictators. The hybrid design allows us to interpret this difference to be driven by the lack of any element of reciprocation in the choice faced by a dictator.

4. Discussion

The interpretation of data from lab experiments may be difficult if there is a mismatch between the experimenter's and subjects' models of the interactions. The between-subjects implementation is probably the unwritten norm for most classes of experiments in economics (Camerer, 2003, pp. 41). The within-subjects implementation prioritizes reducing differences in behavior *across* different games that could arise in the between-subjects implementation because of different subjects playing the different games. The hybrid implementation prioritizes reducing the mismatch in perceptions of the same game that could arise in the between-subjects implementation because of the abstract nature of a game.[13]

In contrast with the results obtained by Dufwenberg and Gneezy (2000) using a between-subjects design, the results we obtain using the hybrid implementation of trust and dictator games reveal a significant difference in the behavior of trustees and dictators.[14] It seems that information about the trust game provides a reference for dictators and information about the dictator game provides a reference for trustees, something which is not possible in the between-subjects implementation. This interpretation is further supported by the pattern in the beliefs of subjects. A significant majority of subjects perceive the nature of the choice

[13]The data obtained from a hybrid implementation of any given experiment may or may not differ from a standard implementation (within or between). Of course, it is possible in some cases that the interpretation could differ even if the data obtained from a hybrid implementation are identical to the data obtained from, say, the between-subjects implementation.

[14]Needless to remark, differences in subject pools may partly contribute to these differences in the data.

faced by trustees to be different from that faced by dictators even though the trustees made their choices without knowing the choice of the trustor (which differs from the conjecture by McCabe et al., 2003). Given the results of Cason and Casari (2009), we would expect that our results would be further strengthened if trustees made their choice after observing the choice of the trustors.

5. Conclusion

In this paper we have focused on one of the various dimensions of the complex issue of *context* as it relates to the conduct of lab experiments and the interpretation of the resulting data: how to increase the likelihood that subjects attach the same meanings to actions as does the experimenter. The importance of context in studies of social preferences is succinctly expressed by Wilson (2010, pp. 81) when he states that "by their sociality people are trained in conventions to apply, and not necessarily deliberately so, different rules of behavior depending upon the contextual cues of the interaction." The hybrid design proposed in the paper may serve as a useful complement to the between and within subjects designs, especially in studies of social preferences. After all, it is the experience with 'sociality' upon which the experimenter relies while designing an experiment related to social preferences. The hybrid design implicitly exploits this feature and attempts to provide the informational setting where subjects can use their experiences with the same 'sociality' to arrive at the meanings for actions that the experimenter believes in.

The hybrid design differs from the notions of 'framing' and 'priming' in a fundamental manner. After identifying the treatments to test the relevant hypotheses, the experimenter can frame the experiment or prime the subjects in several ways. In contrast, there will exist only one way to conduct the hybrid implementation of the experiment. The hybrid design may be thought of as a step towards finding the appropriate middle ground between the polar views of presenting experimental instructions in abstract terms or framing them with details of a particular context in order to enhance interpretability of results without compromising much with their generality.

We conclude by noting that the discussion in the present paper would be vacuous if there existed designs and ways to deliver the instructions that help subjects to perfectly understand the abstract treatment condition where they make their decision. In other words, subjects do not necessarily need to interpret the choices available to them in light of alternative treatments that they could participate in. The ultimate objective is to reduce the mismatch between subjects' and experimenter's models of any interaction and all ways of achieving this need to be explored.

References

[] Ashraf, N., Bohnet, I., and Piankov, N. (2206). "Decomposing Trust and Trustworthiness." *Experimental Economics*, 9, pp.193-208.

[] Bardsley, N. (2008). "Dictator Game Giving: Altruism or Artefact?" *Experimental Economics*, 11, pp. 122-133.

[] Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C. and Sugden, R. (2010). *Experimental Economics: Rethinking the Rules.* Princeton University Press.

[] Berg, J., Dickhaut, J., and McCabe, K. (1995). "Trust, Reciprocity and Social History." *Games and Economic Behavior*, 10, pp. 122-42.

[] Birnbaum, M. H. (1999). "How to Show that 9 > 221: Collect Judgments in a Between-subjects Design." *Psychological Methods*, 4, pp. 243-249.

[] Blanco, M., Engelmann, D., and Norman, H. T. (2008). "A Within-Subject Analysis of Other-Regarding Preferences." *Games and Economic Behavior*, Forthcoming.

[] Bohnet, I., Fiona, G., Benedikt, H., and Zeckhauser, R. (2008). "Betrayal Aversion." *American Economic Review* 98(1), 2008, pp. 294-310.

[] Bolton, G. E., Brandts, J., and Ockenfels, A. (1998). "Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game." *Experimental Economics*, 1(3), pp. 207-219.

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

[] Bolton, G. E. (2010). "Testing Models and Internalizing Context." *Journal of Economic Behavior and Organization*, 73, pp. 16-20.

[] Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton, NJ: Princeton Univ. Press, 2003.

[] Casari, M. and Cason, T. N. (2009). "The Strategy Method Lowers Measured Trustworthy Behavior." *Economics Letters*, 103, pp. 157-159.

[] Cherry, T. D., Frykblom, P., and Shogren, J. F. (2002). "Hardnose the Dictator." *American Economic Review*, 92(4), pp. 1218-1221.

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

Modeling Economic Behavior." *Experimental Economics*, 11, pp. 1-24.

[] Cox, J. C. (2010). "Some Issues of Methods, Theories, and Experimental Designs." *Journal of Economic Behavior and Organization*, 73, pp. 24-28.

[] Dana, J., Cain, D. M., and Dawes, R. M. (2006). "What You Don't Know Won't Hurt Me." *Organizational Behavior and Human Decision Processes*, 100(2), pp. 193-206.

[] Dufwenberg, M. and Gneezy, U. (2000). "Measuring Beliefs in an Experimental Lost Wallet Game." *Games and Economic Behavior*, 30(2), pp. 163-182.

[] Dyer, D. and Kagel, J. H. (1996). "Bidding in Common Value Auctions: How the Commercial Construction Industry Corrects for the Winner's Curse." *Management Science*, 42(10), pp. 1463-1475.

[] Eckel, C. C. and Wilson, R. K. (2004). "Is Trust a Risky Decision?" *Journal of Economic Behavior and Organization*, 55(4), pp. 447-465.

[] Ellingsen, T. and Johannesson, M. (2008). "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review*, 98(3), pp. 990-1008.

[] Engelmann, D. and Strobel, M. (2000). "The False Consensus Effect Disappears if Representative Information and Incentives Are Given." *Experimental Economics*, 3, pp. 241-260.

[] Fehr, E. and Schmidt, K. M. (1999). "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3), pp. 817-868.

[] Harrison, G. W. and List, J. A. (2004). "Field Experiments." *Journal of Economic Literature*, 42(4), pp. 1009-1055.

[] Hoffman, E., McCabe, K., and Smith, V. L. (1996). "Social Distance and Other-regarding Behavior in Dictator Games." *American Economic Review*, 86(3), 653-659.

[] Kahneman, D. (2003). "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review*, 93(5), pp. 1449-1475.

[] Lambdin, C. and Shaffer, V. A. (2009). "Are within-subjects designs transparent?" *Judgment and Decision Making*, 4(7), pp. 554-566.

[] Levine, D. K. (1998). "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics*, 1(3), pp. 593-622.

[] Levitt, S. D. and List, J. A. (2007). "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives*, 21(2), pp. 153-174.

[] Levitt, S. D., List, J. A, and Reiley, D. H. (2010). "What Happens in the Field Stays in the Field." *Econometrica*, 78(4), pp. 1413-1434.

[] List, J. A. (2007). "On the Interpretation of Giving in Dictator Games." *Journal of Political Economy*, 115(3), pp. 482-493.

[] McCabe, K., Rigdon, M. L., and Smith, V. L. (2003). "Positive Reciprocity and Intentions in Trust Games." *Journal of Economic Behavior and Organization*, 52(2), pp. 267-275.

[] Orne, M. T. (1962). "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and their Implications." *American Psychologist*, 17(11), pp. 776-783.

[] Orne, M. T. (1973). "Communication by the Total Experimental Situation." *Communication and Affect.* Second Edition. Eds. P. Pliner, L. Krames and T. Alloway. Academic Press.

[] Ortmann, A. and Gigerenzer, G. (1997). "Reasoning in Economics and Psychology: Why Social Context Matters." *Journal of Institutional and Theoretical Economics*, 153(4), pp. 700-710.

[] Samuelson, L. (2005). "Economic Theory and Experimental Economics." *Journal of Economic Literature*, 43(1), pp. 65-107.

[] Smith, V. L. (2003). "Constructivist and Ecological Rationality in Economics." *American Economic Review*, 93(3), pp. 465-508.

[] Smith, V. L. (2010). "Theory and Experiment: What are the Questions?" *Journal of Economic Behavior and Organization*, 73, pp. 3-15.

[] Sobel, J. (2005). "Interdependent Preferences and Reciprocity." *Journal of Economic Literature*, Vol. 43(2), pp. 392-436.

[] Tversky, A., and Kahneman, D. (1981). "The framing of Decisions and the Psychology of Choice." *Science*, 211, pp. 453-458.

[] Wilson, B. J. (2008). "Language Games of Reciprocity." *Journal of Economic Behavior and Organization*, 68, pp. 365-377.

[] Wilson, B. J. (2010). "Social Preferences aren't Preferences." *Journal of Economic Behavior and Organization*, 73, pp. 77-82.

[] Zizzo, D. J. (2010). "Experimenter Demand Effects in Economic Experiments." *Experimental Economics*, 13(1), pp. 75-98.

Appendix I: Beliefs

Table 11 provides a summary of the beliefs of trustors in the trust game and recipients in the dictator game. Table 12 separately summarizes the beliefs of trustors who chose IN and those who chose OUT. Table 13 provides a summary of the beliefs of trustees in the trust game and dictators in the dictator game. Table 14 separately summarizes the beliefs of trustees who gave strictly more than the outside option to trustors and those who did not.

Table 11: Beliefs of trustors and recipients

	Trustors in Game T^*				**Recipients in Game D^***			
	F_{AB}	S_{ABA}	F_{AA}	F_{AD}	F_{CD}	F_{CA}	F_{CB}	S_{CAB}
Median	5.00	4.00	3.00	5.00	4.00	4.00	5.00	6.00
Mean	5.65	4.20	3.48	4.90	4.65	3.73	5.58	5.68
Std. Dev.	3.53	2.21	2.17	3.25	3.56	2.10	3.19	3.39

Table 12: Beliefs of trustors conditioned on choice

	Trustors who chose IN				**Trustors who chose OUT**			
	F_{A_IB}	S_{A_IBA}	F_{A_IA}	F_{A_ID}	F_{A_oB}	S_{A_oBA}	F_{A_oA}	F_{A_oD}
Median	9.00	5.00	5.00	6.00	3.00	4.00	2.00	5.00
Mean	7.74	5.05	4.74	5.90	3.76	3.43	2.33	4.00
Std. Dev.	3.30	1.54	1.63	3.43	2.59	2.46	1.98	2.86

Table 13: Beliefs of trustees and dictators

	Trustees in T^*				**Dictators in D^***			
	F_{BA}	S_{BAB}	F_{BB}	F_{BD}	S_{DCD}	F_{DD}	F_{DA}	F_{DB}
Median	4.50	8.00	5.00	4.00	5.00	4.00	3.00	5.00
Mean	4.50	7.42	5.72	4.83	5.42	4.50	3.50	5.15
Std. Dev.	2.11	3.73	3.68	3.19	3.59	3.62	2.12	3.11

Table 14: Beliefs of trustees conditional on choice

	Trustees who gave more than 5				**Trustees who gave at most 5**			
	F_{B_HA}	S_{B_HAB}	F_{B_HB}	F_{B_HD}	F_{B_LA}	S_{B_LAB}	F_{B_LB}	F_{B_LD}
Median	5.00	10.0	8.00	8.00	4.00	5.00	3.00	3.00
Mean	5.15	8.80	8.00	6.75	3.85	6.05	3.45	2.90
Std. Dev.	1.76	2.07	3.03	2.99	2.28	4.50	2.78	2.02