

Kim, Jeongbin; Putterman, Louis G.; Zhang, Xinyi

Working Paper

Trust, beliefs and cooperation: An experiment

Working Paper, No. 2019-3

Provided in Cooperation with:

Department of Economics, Brown University

Suggested Citation: Kim, Jeongbin; Putterman, Louis G.; Zhang, Xinyi (2019) : Trust, beliefs and cooperation: An experiment, Working Paper, No. 2019-3, Brown University, Department of Economics, Providence, RI

This Version is available at:

<https://hdl.handle.net/10419/202609>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Trust, Beliefs and Cooperation: An Experiment

Jeongbin Kim, Louis Putterman and Xinyi Zhang*

March 2019

Abstract: We use a two-phase experimental design to study how systematically manipulated beliefs about trust and trustworthiness can promote or deter cooperation. We use decisions in an initially played trust game to create five environments that differ in the information subjects have about the relative trust/trustworthiness of fellow group members when they make a voluntary contribution decision in our experiment's second phase. We find that perceived high trusting environments are treated equivalently to ones of perceived high trustworthiness, with both positively affecting subjects' first-order beliefs about the cooperativeness of group-mates, and in consequence, leading to higher contributions. Our results thus suggest that people cooperate more and hence produce more together in an environment of high trust/trustworthiness, indicating one channel through which trust helps to grow the economic pie.

* The authors are, respectively, post-doctoral scholar, California Institute of Technology; professor of economics, Brown University; and Ph.D. candidate in economics, Brown University. We are grateful to Brown University for providing funding for this experiment. We thank Pedro Dal Bo, Geoffroy de Clippel, Xu Zhang, and other participants in the theory workshop at Brown University for helpful comments and suggestions.

1. Introduction

Trust has been regarded as an important influence upon, or at least correlate of, various aspects of economic prosperity, including the rate of economic growth (Knack and Keefer, 1997; La Porta et al., 1997, Algan and Cahuc, 2014). Zak and Knack (2001) presented influential evidence that trust affects growth, offering the interpretation that trust is a feature of the social, economic and institutional environments in which economic transactions take place, and that higher trust reduces transaction costs, which in turn engenders a higher investment rate and faster economic growth. Their conclusion echoes Arrow's (1972) argument that "Virtually every commercial transaction has within itself an element of trust, ..., much of the economic backwardness in the world can be explained by the lack of mutual confidence." The underlying idea is that a prerequisite for the successful development of market economies is to enlarge the scope of interactions to include anonymous others, and that not all risks of interacting with others can be removed by legal rules and sanctions (Algan and Cahuc, 2010; Fukuyama, 1995).

Trust and trustworthiness are widely operationalized by economists in terms of sequential interactions in which a first mover "make [themselves] vulnerable to others' actions" (Fehr, 2009) before the latter respond with more or with less trustworthiness. However, cooperation in more simultaneous and symmetric dilemmas of collective action is also important to a well-functioning economy (Ostrom, 2010), and may also be influenced by trust and its absence. In many situations where self-interest might otherwise lead to free-riding, cooperation is a key to enhancing efficiency. Examples include voluntary provision of local public goods (Ledyard, 1995), cooperation among partners or workers of enterprises employing profit-sharing schemes (Bonin, Jones and Putterman, 1993), and efforts to establish and maintain institutions with less theft and corruption (Tabellini, 2010). Cooperation in these domains is an important contributor to overall economic efficiency and thus growth.¹

However, an empirical question that remains to be answered is exactly what mechanisms lie behind the effect of trust on cooperation. One plausible story for explaining the associations between trust and cooperative outcomes focuses on beliefs: people cooperate because they believe others will also cooperate and/or that others have expectations of high cooperation among those whom they encounter. In other words, members of a society with high trust may share optimistic beliefs about others' behaviors or beliefs, and this may lead them to be more willing to contribute their effort in cooperative endeavors. Of course, this explanation makes sense only if many people prefer to cooperate when they believe others also do so, a preference that must over-ride material self-interest and that is identified in the literature as conditional cooperation (Keser and Van Winden, 2000; Fischbacher, Gächter and Fehr, 2001). Our study joins others (Thöni and Volk, 2018) in finding considerable evidence of its presence.

A problem of the approach just sketched, however, is that it is hard to identify the effect of beliefs on cooperation in natural settings, since societies or groups have been formed endogenously, and what people believe is difficult to know even if survey responses are available.

¹ There has been a wide range of related discussion in the literature. For instance, see Sapienza, Toldra and Zingales (2013).

Although many papers posit the importance of trust by highlighting its effects on or at least correlation with economic growth, proving specific mechanisms by which trust promotes economic activity, including cooperation, can be difficult with observational data. It is unclear, for instance, whether beliefs lead to cooperation, or whether causality runs in the opposite direction.

In this paper we present a laboratory experiment to shed light on the roles of trust, trustworthiness and beliefs as channels through which cooperation among economic actors can be promoted or deterred. Subjects are first asked to play a trust (also called investment) game in both roles – that of first and second mover. Then, they move to a second phase in which they participate in a voluntary contribution mechanism (hereafter VCM, referred to as linear public goods game by some authors). In order to manipulate beliefs about the other members in a group, we use trust game behavior as the basis for generating five environments with different levels of laboratory-measured trust and laboratory-measured trustworthiness: a group in which people are randomly matched, groups in which the average level of trust is relatively high (low), and groups in which the average level of trustworthiness is relatively high (low), respectively. Although we describe only relative and not absolute behaviors, leaving open the possibility that ranking by trust and trustworthiness is entirely random, our subjects (correctly) assume that behaviors vary, an interesting finding in its own right. Each subject plays a one-shot VCM consecutively and without feedback in each of the five environments (groups), and their first-order and second-order beliefs about contributions in each group are obtained by an incentivized elicitation. Play follows the strategy method (Selten, 1965) in that subjects know that only one randomly chosen environment will be the basis for their payoff.

The two games are chosen with careful consideration of what we can infer from subjects' behavior. We use a trust game in the first phase because there is evidence that behavior as the first and second mover in this game can capture inclination to be trustworthy, fair, or reciprocating, as well as reflecting beliefs about such inclinations in others, which are importantly influenced by own type via introspection. The relationship between the incentivized experimental and survey measures of trust is discussed in Section 2. We employ the VCM in the second phase because it presents a multi-person social dilemma which resembles many situations in the real world where full cooperation leads to efficiency. While both the trust game and the VCM constitute social dilemmas, the asymmetric and sequential nature of the first versus the symmetric and simultaneous nature of the second game are contrasting features on which we provide a fresh perspective in the next section.

Having elicited for each environment subjects' first-order beliefs about others' inclinations to cooperate, as well as their second-order beliefs about others' first-order beliefs, we show that subjects positively associate both trust and trustworthiness with cooperation, and that they are approximately equally more cooperative when in a high trusting as when in a highly trustworthy environment. By looking at the effects of first-order and of second-order beliefs on cooperation separately and simultaneously, we also find that the effects of the first-order beliefs outweigh those of the second-order ones. Finally, we show that subjects who sent (returned) more in the trust game are more likely to be conditional cooperators or altruists in the VCM. These findings imply that

reciprocity and beliefs about others' reciprocity are key channels leading to the higher level of cooperation in more trusting and trustworthy environments.

Our results may have important implications outside the laboratory. We identify a channel through which trust and trustworthiness, each representing a potentially distinct behavior, can affect the level of cooperation through affecting beliefs about others' likelihood of cooperating. This implies that while social capital in general is something that should be enhanced or safeguarded whenever possible, establishing a foundation for belief that others will not exploit one's own vulnerability may be especially important insofar as economic growth depends in part on cooperative effort. Institutions that can be counted upon to punish the more egregious and identifiable cases of exploitation of trust can make it rewarding to invest in fostering social preferences within families and in other settings, and people with social preferences can sustain good institutions, a virtuous circle.

The structure of our analysis is as follows. We review related literature in Section 2. Section 3 explains our experimental design in detail. We present the theoretical background and behavioral hypotheses in Section 4. Section 5 shows our analyses and results. Section 6 discusses some caveats and concludes.

2. Related literature

2.1. Measuring trust

The attempt to measure trust and to study how trust is related to economic activity, institutions, and growth has been an active area of research by economists for more than two decades. Early studies used as their measure of trust the answers to the standard World Values Survey (WVS) trust question ("Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?"). Knack and Keefer (1997) concluded that higher trust, thus measured, is conducive to growth based on cross-country evidence for a sample of 29 market economies. La Porta *et al.* (1997) found similar evidence of a correlation between trust and civic participation, and Guiso *et al.* (2009) found that trust is positively related at country level to the volume of international trade. Other survey-based measures, such as responses to a question about the likelihood of a lost wallet and its contents being returned, have been used in a similar fashion.

At the same time, vagueness and lack of agreement about what the survey measures of trust truly capture helped fuel interest in an initially separate stream of research in the experimental economics literature. Berg, Dickhaut, and McCabe (1995) had introduced an "investment game" in which the first to move, of two players, can be made better off by transaction with the second, but only if the latter responds to an unsecured transfer from the former—tripled by the experimenter—in a reciprocating manner, and if the former thus makes that transfer in the presumed anticipation of such trustworthiness, an action or state of mind that many researchers subsequently denoted trust (e.g., Fehr, 2009; Eckel and Wilson, 2011). Assuming that preferences such as altruism are not the important motivations behind the act of the first mover (Cox, 2004; Ashraf *et al.*, 2006), an assumption that the Berg *et al.* design renders plausible by recruiting first

and second movers from a common subject pool and endowing them with equal initial endowments, the trust interpretation is a natural one.² The central element of the trust decision is the tradeoff between exposing oneself to the risk of being “exploited” or “betrayed” (Bohnet and Zeckhauser, 2004; Ben-Ner and Putterman, 2001) and the possibility of achieving higher-payoff outcomes (Thöni, 2015).

A number of researchers have investigated how the survey measures of trust are related to specific behavior in the incentivized experimental trust game, reaching differing conclusions. Glaeser *et al.* (2000) and Lazzarini *et al.* (2005) found that the answers to the WVS trust question are not significantly correlated with first movers’ sending behaviors, but are related to second movers’ returning behaviors, in the trust game. But Fehr and Fischbacher (2003) and Bellemare and Kroeger (2007) obtain opposite results, i.e. first movers’ sending behaviors correlate with the answers to WVS-like questions, in their findings, and similar correlations are found by Falk *et al.* (2016) and Murtin *et al.* (2018).³ The Fehr and Fischbacher and Bellemare and Kroeger papers fail to find significant correlations between responses to their survey trust questions and second movers’ returning decisions. The contradictory findings suggest that survey-measured trust could be significantly correlated with both lab-measured trust and lab-measured trustworthiness, which seems to imply that not only do trust and trustworthiness, as measured in the game, appear to be non-separable (Fehr, 2009), but people tend not to distinguish them in real life. Such “non-separability” comes from the facts that beliefs in trustworthiness of others plays a significant role in explaining why sending varies (Thöni *et al.*, 2012) and that players extrapolate their opponent’s behavior from their own (Sapienza *et al.*, 2013): the belief in the trustworthiness of others is often correlated with one’s own trustworthiness since it is obtained partly by introspection. Nevertheless, the fact that lab-measured trust and trustworthiness sometimes correlate positively to WVS results suggests a linkage between laboratory findings and real world implications.

Although first mover sending in the trust game may be closely linked to the second mover returning tendency through the channel of beliefs, it is nevertheless important to be clear about the asymmetry of the two decisions, from the standpoint of economic theory. Put simply, faced with an environment in which most people are trustworthy, the decision to send money as first mover can be fully explained by self-interest and rationality, whereas this is never the case for returning money as second mover in the absence of reputational considerations (which the standard experimental design rules out). On the other hand, as considered next, second mover returning

² We abstract from atypical cases in which the first mover is informed of special neediness on the part of the recipient, for example the microfinance lending case studied by Chen *et al.* (2017). The interpretation of first mover sending as trusting also requires assuming absence of strong efficiency motives, i.e. the first mover does not strongly value making the pie larger irrespective of who consumes it, a motive for which Charness and Rabin (2002) find some evidence, but one that seems unlikely to be of first order importance in environments like that of Berg *et al.* (1995).

³ Falk *et al.* (2016) report the statistical testing on which Falk *et al.* (2018) base their use of two proxies for trust—a non-monetized survey response choice in a hypothetical trust game, and a survey trust question—as a measure of trust, identified with first mover choices in an incentivized laboratory trust game. Murtin *et al.* (2018) find a statistically significant correlation between responses to two survey trust questions including the WVS one, on the one hand, and first mover choices in an incentivized trust game, in a representative survey instrument being tested in an OECD project denoted TrustLab in six countries, although they find other experimental choices, including ones in a dictator game, closely correlate with first mover trusting.

resembles contributing money in the one-shot VCM in that its explanation requires preferences additional to self-interest, and preferences of much the same kind can be at work in each case.

2.2. *Trust and Conditional Cooperation in the VCM*

The VCM is a much-studied experimental game, designed to investigate voluntary cooperation in the provision of a public good (Isaac *et al.*, 1984, Ledyard, 1995, Zelmer, 2003, Chaudhuri, 2011). It holds special interest for us first because cooperation in the sense of adherence to norms opposing corruption, theft, and nepotism, as well as cooperation in partnerships and in the workplace, are among the drivers of economic prosperity and growth. Demonstrating an impact of trust (proxied by trust game behaviors) on cooperation (represented by VCM choices) would therefore constitute evidence of a channel through which trust can enhance GDP or its growth. Second, while both the VCM and the trust game model dilemmas in which social efficiency is in conflict with strict self-interest, the one-shot VCM is a simultaneous and symmetric social dilemma that from a certain standpoint collapses the decisions of the Berg *et al.* (1995) trustor and trustee into a single choice (Thöni, 2017). Although a self-interested individual will never contribute to the group project in the VCM, research beginning with Fischbacher, Gächter and Fehr (2001) has suggested that the modal behavioral type in this game is in fact conditionally cooperative, meaning that the player prefers to contribute provided that others do so,⁴ with contributing being explicable as positive reciprocity (Rabin, 1993, Hoffman, McCabe and Smith, 1998, Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006) or as inequality aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000).⁵ Since reciprocity and inequality aversion can also provide explanations of second movers' decisions to return money in the trust game, we hypothesize that second movers' returning choices in the trust game and the same individuals' contribution choices in the VCM will be highly correlated, especially if we control for beliefs about others' contributions. Observing how our subjects act as both trust game first and second movers and as VCM decision-makers from whom beliefs about others' actions are elicited can provide us with new insights into the motivational underpinnings of pro-sociality in these important and distinctive environments.

Our paper also shares some features with previous experimental research that has provided further evidence of conditional cooperation by exogenously or endogenously manipulating group composition and observing that high contributors to a public good tend to continue to make high contributions if interacting primarily with other high contributors. An example is multi-period VCM experiments in which participants are matched by the experimenter with others of like

⁴ Replications using the method first introduced by those authors include Kocher *et al.* (2015), Fischbacher and Gächter (2010), Putterman *et al.* (2011), and Thöni *et al.* (2012).

⁵ Akerlof (1982) references the sociological literature's observation of reciprocity as a norm underlying much of human behavior. The closely related preference for cooperating provided that the counterpart cooperates, despite higher monetary payoff for choosing defection under that premise, is discussed by Sen (1967) as a case where an agent has preferences rendering the problem one of "assurance," despite facing the monetary payoffs of a prisoners' dilemma. As Gintis (2009) and others state, the preferences associated with Sen's assurance problem make the game in subjective payoffs a stag hunt game.

disposition, including Gächter and Thöni (2005), Gunnthorsdottir *et al.* (2007), and Ones and Putterman (2007), which show that when high contributors are grouped by the experimenter with other high contributors, their contributions in the VCM are sustained at high levels. Similar findings arise in the endogenous matching experiment of Page *et al.* (2005), in which subjects rank preferred partners in a VCM and have an algorithm assign them to groups according to mutual preference; here, like contributors become sorted by the mechanism and higher contributors maintain that behavior in part due to observing that others do the same. Apart from conditional cooperation, these experiments suggest the importance of beliefs, and their designs can be seen as embodying belief formation devices, although unlike the present paper, they operate within a single game form and do not conduct belief elicitations.

Several papers have investigated whether survey-measured or lab-measured trust, or both, are associated with contribution behavior in the VCM. Using 630 subjects in rural and urban Russia, Gächter *et al.* (2004) found that whereas answers to the WVS trust question are not correlated with behavior in a one-shot VCM, subjects who respond to another WVS question that most others are fair or helpful are more likely to contribute in the VCM. Thöni *et al.* (2012) delve into this problem in depth by using a representative sample in Denmark. Subjects in their study are asked to play a VCM that elicits separately both their conditional cooperation in strategy method decisions and their beliefs about others' (unconditional) cooperation. They find that responses to the trust questions have strong correlations with conditional cooperativeness. Kocher *et al.* (2015) study a design in which subjects engage first in a trust game, then a strategy method VCM à la Fischbacher *et al.* (2001). They show that in general, first-mover trust correlates positively with the unconditional contribution and negatively with the propensity to be a free rider rather than a conditional cooperator in the VCM.

Our paper adds to the research just mentioned by adding new observations of how trust and trustworthiness correlate with cooperative behavior in the VCM, and by strengthening understanding of what lies behind these correlations.⁶ Our main contribution, however, lies in showing how impressions of the trust and trustworthiness of others, impressions based partly on information conveyed by us but also crucially on the beliefs subjects bring to the lab with regard to how these characteristics vary within the population, can also shape beliefs about others' cooperativeness, and (second order) beliefs about others' beliefs about the cooperativeness of others, and how these beliefs affect their costly decisions to cooperate or not when facing different sets of counterparts. We thereby illuminate a specific channel through which trust and trusting encourage cooperation.

2.3. Manipulating beliefs

⁶ Papers that investigate cross-game behavior associations with other games include Dariel and Nikiforakis (2014), who perform a within-subject analysis in a VCM and a gift-exchange game, finding that conditional cooperators tend to reciprocate higher wages in the gift-exchange game with high levels of effort, while non-cooperators do not exhibit this tendency. Another example is Blanco *et al.* (2011), who have subjects play four different types of games, including a VCM, to test for cross-game evidence of inequality aversion, finding strong predictive power for measures of this preference.

Our paper is related also to other studies that manipulate beliefs in strategic interactions. Charness and Dufwenberg (2006) report a correlation between behavior and second-order beliefs, which may be affected by the second mover's pre-play communication in a modified trust game. To avoid consensus effects and achieve exogenous variations in beliefs, Dufwenberg *et al.* (2011) use game "framing." In particular, exploiting a well-known result in psychology by Liberman *et al.* (2004), they show that depending on the framing of a VCM by Community game or Market game labeling, subjects have systematically different first- and second-order beliefs about others' contributions and expectations, which lead to changes in own cooperation. They find evidence for what they call "guilt aversion," that is a desire to act pro-socially or favorably towards others primarily because one believes that they expect this of you. Ellingsen *et al.* (2010) try to manipulate subjects' second-order beliefs by disclosing the first-order beliefs of a person whom they are paired with. A recent paper by Khalmetski (2016) manipulates second-order belief by changing the probability of a game being played, where the true state of the world is only known to the sender of a message. Our paper contributes to this stream of literature by introducing another way of manipulating beliefs, namely providing information about other group members' relative behaviors in a previously played game. Our results contribute to the literature on guilt aversion insofar as they contrast with those of Dufwenberg *et al.* (2011), i.e., in our setting second order beliefs are not found to be a significant determinant of cooperation once first order beliefs are controlled for.

3. Experimental Design and Procedure

Our experiment consists of two phases: the first phase involves a trust game, and the second phase, a voluntary contribution mechanism (hereafter VCM) and associated belief elicitations, in different environments (groups). Decisions in the first phase determine group composition in most environments in the second phase. Nonetheless, the two phases are independent in the sense that the instructions in the first phase give no hint of its importance for the second, there is no feedback from that phase's decisions before the second phase begins, and instructions for the second phase are distributed only after the end of the first phase, to avoid strategic response and contamination of the first phase decisions. Decisions in both phases are incentivized and monetarily rewarded only at the end of the session.

Everything described in the following sub-sections was common knowledge among all subjects. The instructions are included in the online Appendix. Our experiment consists of a single treatment with 120 participants in six sessions making decisions for three roles (trust game first mover and second mover and VCM contributor in five potentially payoff-determining groups) as we now detail.

3.1. The Trust Game (First Phase)

We used a slightly modified version of the original trust game designed by Berg *et al.* (1995), strictly parallel to the original in terms of decisions and payout structure, but differing in that each subject made decisions as both sender and receiver, with the latter decisions taken by strategy

method. After reading instructions explaining this structure, subjects first made decisions as first mover (sender), then made decisions as second mover for each possible amount that might be received. They were told that the two decisions are independent in the sense that for purposes of payment they would be randomly assigned to one or the other role and matched with a randomly chosen counterpart who is (also randomly) assigned the opposite role. Rather than physically divide ten one dollar bills as in Berg *et al.*, subjects, as in most subsequent experiments, entered their choices in the computer. Like Berg *et al.*, we gave both first and second mover equal endowments, these being in our case 50 tokens, of which the first mover could send any multiple of 5, yielding eleven options as in the original experiment. Also as in Berg *et al.*, any sent amount was tripled, and the receiver could send back any integer amount between 0 and the tripled amount (the second mover at a minimum kept her endowment). Returning decisions as receiver were made in a contingency table, conditional on each of the 10 possible positive received amounts (there being no decision to make in the case of being sent 0). As is standard, the returned amount was not tripled.

To summarize more formally, payoffs for subject i as 1st mover (sender) can be written

$$\pi_i = 50 - a + b \quad (1),$$

and for subject j as 2nd mover:

$$\pi_j = 50 + 3a - b \quad (2),$$

where $a \in [0, 5, \dots, 50]$, and $b \in [0, 1, 2, \dots, 3a]$.

In the trust game, the sub-game perfect equilibrium when sender knows receiver to be a rational maximizer of own payoff, is for the selfishly rational sender to send nothing, with the result that both simply keep their 50 token endowments. Pareto improvement is possible if the first mover sends a positive amount and the second mover returns at least 1/3 of what she receives; for example, both can end up with a doubled amount, 100 tokens, if the first mover sends his full endowment and the second mover returns 2/3 of the received 150 tokens. In practice, the sending behavior in the game is affected by the sender's beliefs about the recipient's likelihoods of returning various amounts, how negatively the sender weighs potentially negative outcomes (a.k.a. "betrayal aversion"), and perhaps other preferences (Sapienza *et al.*, 2013). In our discussion, we treat the amount sent as a measure of trusting, as in much of the literature (Eckel and Wilson, 2011; Johnson and Mislin, 2011). Returning money in the one-shot trust game can never be explained by self-interest. We further discuss motivation of this behavior, which we call trustworthiness, in Section 4.

3.2. The Voluntary Contribution Mechanism and Belief Elicitations (Second Phase)

In the second phase, each subject decided simultaneously, for five different environments differing in information on group membership, how to allocate a 20 token endowment between a private

and a group account in a one-shot linear VCM with group size 5. On successive screens, each also reported first and second order beliefs for each environment.⁷ After subjects made their contribution decision for each of the five environments, that is, they were prompted to provide their first-order beliefs about what the others in that group would contribute, on average, in each environment. Finally, they were asked to indicate their second-order beliefs, that is, what they believed the other members of each group would on average list as their own first-order belief about how much group members other than themselves would allocate to the group account. We asked subjects to make all contribution decisions, then state all beliefs rather than to make a contribution and state associated beliefs environment by environment, in order to prevent possible contamination of contribution choices by previously elicited beliefs. Subjects were truthfully told that one of the five environments would be randomly selected, in the end, to determine payoffs for this phase.⁸ To incentivize truthful estimates, following Dufwenberg *et al.* (2011), a subject was given five additional tokens of earnings if her first-order belief for the selected environment was within one token of the true average contribution of the other four group members, and likewise if her second order belief was within one token of the true average of the first-order estimates of the other four group members.⁹ The total earnings in this phase were thus the earning from the contribution decision plus any rewarded amount from the estimates.

As mentioned above, the five environments differed with respect to prospective group composition, and the information given to the subjects about it. In particular, groups in environment 1 were to be formed randomly from among all session participants, while group memberships in the other four environments was formed with the aid of rankings of first phase behaviors. Put somewhat too simply, these groups' member sets were (on average) 'relatively trusting,' 'relatively untrusting,' 'relatively trustworthy,' and 'relatively untrustworthy' experiment participants, respectively. More precisely, in each session, we ranked all trust game sending decisions, and all returning decisions for the contingency of receiving the highest possible amount, and assigned the corresponding subjects numerical ranks 1 – 20 for (lowest to highest) first mover sending, and (separately) 1 – 20 for (lowest to highest) second mover return proportion, with ties broken randomly. This way, each subject was identified by two numbers, one from her sending decision reflecting her "trust" rank, and the other from her returning decision reflecting her "trustworthiness" rank. Because ties were broken randomly, these ranks would be informative only to the degree that subjects differed in their Phase 1 decisions. Subjects were truthfully told that in environment 2, the average rank of the other group members for first mover sending is above 12; in environment 3, the average rank for first mover sending is below 8; in environment

⁷ The experimental screens are included in the online Appendix.

⁸ If an environment was selected for payout of one participant, then the other four group members in that same environment were paid according to their decisions in it, so there is no deception regarding the impact of one's decision on not only oneself but a definite set of other participants.

⁹ A potential criticism of our incentivization scheme is that it would allow subjects to hedge by providing a guess as close to the theoretically predicted average as possible (Blanco *et al.*, 2010), even though we try our best to control for the timing of elicitation to avoid contamination. A more ideal procedure might be to adopt some kind of "scoring rules" (see the literature survey by Schotter and Trevino, 2014) to make it a dominant strategy to reveal beliefs truthfully. We believe that our design offers subjects little scope for strategic manipulation of beliefs, however, and we think it likely that they simply provide their best guesses of others' choices in the hope of boosting their earnings.

4, the average rank for second mover return proportion (in the highest contingency) is above 12; and in environment 5, that rank is below 8.¹⁰ The terms trust and trustworthy were never used, the instructions mentioning only amounts chosen in specific decisions in each role in Phase 1. Also note that the instructions left open the possibility that all subjects had chosen the same amount as first mover, and likewise that all had chosen the same amount as second mover. Thus, not only were the concepts of trusting and trustworthiness not explicitly invoked, but also suppositions that averaging in the top 12 or bottom 8 implied appreciably different behaviors and tendencies would be strictly the “home grown” beliefs of the subjects themselves, since ranks could have been given entirely randomly.

Table 1 Difference between Environments

Environment	Brief Descriptions (Ranking is in an ascending order: the lowest rank is denoted as 1, and so on)
Random	Matching is done randomly in the computer program.
High Trust	For each participant, the average sending rank of the other four group members is above 12
Low Trust	For each participant, the average sending rank of the other four group members is below 8
High Trustworthiness	For each participant, the average returning rank of the other four group members is above 12
Low Trustworthiness	For each participant, the average returning rank of the other four group members is below 8

Note: environment names shown in Table 1 were not used in experimental instructions or screens.

Table 1 summarizes the five groups or environments for which subjects made VCM contribution decisions and estimates of first and second order beliefs, using convenient environment labels that reflect our conceptual intuitions but that were not used with our participants, to preserve a more neutral framing. In the standard linear VCM, a token allocated by any to the group account yields a payoff of 0.4 to all, and a token allocated to her private account yields a payoff of 1 to her only. The payoff function for any subject i is thus:

$$\pi_i = 20 - c_i + 0.4 * \sum_{j=1}^5 c_j \quad (3)$$

where π_i denotes number of tokens earned by i if the environment is the one paid off on, c_i denotes i 's allocation to the group account, and the summation is over all group members, i included.¹¹ The game is a social dilemma because aggregate payoffs are maximized when all group members choose to contribute all 20 tokens, but each individual's dominant strategy is to free ride and contribute nothing, so Nash equilibrium in the one-shot game entails that no one contributes

¹⁰ It is important to note that subjects were not told about the exact composition of group members in these five environments, but rather were given information about the other four group members' average rank, only. Clearly, it could not always be the case that all five members of a group in which others' trust rank had an average of more than 12 would each be in ranks 13 or higher, since even subjects with low ranks must be capable of being in each environment including this one. Indeed, the lower one's own rank, the higher must be the ranks of the other members on average to assure that for those members, too, the average rank of the other four, including oneself, is 12 or above. Although this means that the expected ranks of other group members could in principle differ by subject, depending on the belief each had regarding her own rank, this rational difference in beliefs by environment is not analytically problematic for us, given that we elicit and control for beliefs.

¹¹ For completeness, we could write π_{ie} , c_{ie} and c_{je} , where e denotes which of the five environments in Table 1 is referenced. Thus, (3) should be understood as a quintuple of potential payoff expressions only one of which is randomly chosen for realization.

anything. We used a different token/money conversion rate for this phase of the experiment, each token earned being converted to \$0.20, as explained in the subject instructions.¹²

3.3. Procedure and Payments

Subjects received copies of instructions for each phase separately at its commencement and were asked to read along as an experimenter read them aloud. The experimenters invited questions and clarified them in private. Subjects had to answer control questions which appeared on their computer screens, to verify their understanding, before commencing play of each phase. At the end of Phase 2, additional beliefs about Phase 1 behaviors, unanticipated because not mentioned in the instructions, were elicited, as described later. An exit survey which included demographic information such as gender, class level, race/ethnicity, and academic major, as well as an unincentivized question about beliefs to which we will refer later, followed. Finally, each subject was shown on the screen which of his/her roles in the first phase was realized, which environment in the second phase was realized, as well as his/her earnings, which averaged \$6.57 for Phase 1, \$6.52 for Phase 2, and \$0.11 for accuracy in belief elicitations. Subjects received their payments in cash in sealed envelopes and then exited the lab.

4. Theoretical framework and hypotheses

4.1. Reciprocity in the trust game and VCM

We propose a simple model which focuses on the role of reciprocity in the trust game and the VCM. People have underlying dispositions towards reciprocity, r . For some people, this disposition is very low and does not impact their behavior. However, for others, this disposition plays an important role both in driving trustworthiness and thus expected trustworthiness and trusting in the trust game, and in driving conditional cooperation and expected cooperativeness in the VCM.¹³ Therefore, in a strategic interaction, one thing that needs more attention is how people form beliefs about others' reciprocity."

We first model decisions in our games in a simple psychological game framework (Geanakoplos *et al.*, 1989), and later discuss its implications for our experimental design.¹⁴ Assume a trust game second-mover j having underlying reciprocity disposition r_j . Ignoring additional factors that might influence trustworthiness, we can simply identify the disposition r of

¹² Because Phase 2 instructions were distributed and read separately from those of Phase 1 and after that phase's completion, it is unlikely that subjects found the difference in token value confusing. We used a higher token value in Phase 2 so that each phase would account for a similar share of payout. Changing the number of tokens available in one phase or the other would have achieved the same effect, but the different endowment sizes and conversion rates could add to subjects' senses of each part as being quite distinct from the other.

¹³ Among other motivations, high dispositions toward reciprocity make people return money if trusted in a trust game, and contribute money if they believe others are contributing in a VCM.

¹⁴ As will be shown later, the main purpose of using a psychological game is to highlight the role of subjects' beliefs about others' reciprocity both in the trust game and in the VCM. Our main qualitative predictions do not hinge on the use of the psychological game. Also, our primary interest is not in fully characterizing the set of equilibria, but in showing existence of an equilibrium where behavior is crucially related to beliefs about others' reciprocity.

a trust game second mover, j , by her reciprocity in that game, denoted by $r_t(r_j)$.¹⁵ Assume that a second mover j 's utility function is defined as follows:

$$U_j = 50 + 3x_i - y_j + \mu_j \min\{y_j - r_t(r_j) \cdot 3x_i, 0\}, \quad (4)$$

where x_i is the amount that first mover i sent, y_j what j sends back, and j 's normatively ideal return rate $0 \leq r_t(r_j) \leq 1$ with $r'_t(r_j) > 0$ is governed by the reciprocity taste parameter, r_j . $\mu_j > 0$ captures the second mover j 's sensitivity to his reciprocity and for simplicity, throughout the paper, it is assumed sufficiently high to render the constraint binding.¹⁶ Assume also r_t to be a known function of the second mover's r_j . $r_t(r_j) \cdot 3x_i$ is, then, j 's benchmark conception of how much she must send back if she is to live up to her ideal of reciprocity, and with $\mu_j > 1$, her utility gain from closing the gap between y_j and that ideal suffices to offset her utility loss from foregoing money payoff y_j .¹⁷ In other words, in a plausible equilibrium that will be discussed below, y_j is a function of r_j . Note that this model is identical to the standard approach if $\mu_j = 0$ or $r_t(r_j) = 0$ for all j .

Next, we model first mover sending. Given the second mover's decision problem, the first mover i maximizes the following utility function,

$$U_i = 50 - x_i + y_j^{ei}, \quad (5)$$

where y_j^{ei} indicates first mover i 's belief or estimate about second mover j 's returning amount, y_j that i cannot know in advance of her decision. Given that y_j is a function of j 's reciprocity r_j , y_j^{ei} can be rewritten as $y_j(r_j^{ei})$, and thus, first mover i 's (first-order) belief about the second mover j 's reciprocity plays a key role in determining the amount sent.¹⁸ Given (5), the value of x_i that maximizes U_i must be one of two values, namely $x_i = 0$ for r_j such that $r_t(r_j^{ei}) < 1/3$, and $x_i = 50$

¹⁵ The following utility function can be regarded as an application of Köszegi and Rabin (2006)'s reference dependence model into other-regarding preferences. Our way of modelling the second mover's behavior is also similar to a model of interdependent preference by Levine (1998) and Gul and Pesendorfer (2016), since in the first game of our experiment, a method is used and second movers make returning decisions for all possible contingent cases of sending behavior. In other words, this rules out strategic uncertainty that second movers may have and allows us to extract information regarding subjects' reciprocity type or personality which will be instrumentally used in the following analysis.

¹⁶ This implies that while some subjects choose $y_j = 0$ in the trust game, it occurs only due to their low reciprocity, not due to their insensitivity to what reciprocity they have. Allowing for low μ_j would increase the number of equilibria but not alter the qualitative aspects of the predictions.

¹⁷ In a more detailed analysis, function r_t might well take into account factors such as the perceived financial neediness of the first mover, how fair the first mover is believed likely to be were he to be in the same position, what the first mover's expectations are, the first mover's likely ethnicity or gender, and so on. But given our focus and the simplicity of our actual experimental design, we can safely leave such details unmodeled.

¹⁸ There may exist other motivations for trusting such as concern for efficiency and adherence to social norms. Having other motivations underlying the first mover behavior in our model does not qualitatively affect the main predictions regarding reciprocity, and for simplicity, we rule out these possibilities.

for r_j such that $r_t(r_j^{ei}) \geq 1/3$. Therefore, in equilibrium, there can be two types of outcomes, one in which there is no sending and returning if $r_t(r_j) < 1/3$, and another in which i sends everything and j returns $y_j = r_t(r_j) \cdot 3x_i \geq x_i$, if $r_t(r_j) \geq 1/3$.¹⁹ Since this strongly bifurcated prediction aligns poorly with existing trust game data, it seems prudent to allow for the possibility that some first-movers are risk averse and form estimates of r_i that assign positive probabilities to a range of values, in consequence of which first mover sending can also take intermediate values. The amount X_i is accordingly, and more generally, assumed to be an increasing function of the first mover's belief about r_j , without necessarily having an abrupt switch point. We omit explicit modification of (5), since the intuition is straightforward and is not fundamental to the relationships on which we focus here.

We next apply the same framework to the analysis of contributing behavior in the VCM. In a group of n members, each group member i maximizes the following utility function:

$$U_i = 20 - c_i + m \sum c_j + \gamma_i \min\{c_i - \frac{r_c(r_i)}{n-1} \sum c_{-i}^{ei}, 0\}, \quad (6)$$

where m is the MPCR, $\gamma_i > 0$ captures i 's sensitivity to his reciprocity and $0 \leq r_c(r_i) \leq 1$ with $r'_c(r_i) \geq 0$. Just like μ_j , we also assume that γ_j takes a high enough value for the constraint to be binding. $\sum c_{-i}^{ei}/(n-1)$ indicates i 's belief or estimate about other group members' average contributions. As shown above, it is straightforward that own contribution c_i is a function of own reciprocity, r_i . In contrast to the case of the trust game, players decide their contributions while they form beliefs about others' reciprocity in the VCM. There can be multiple equilibria in this model. In one extreme case, if $r_i = 0$ for each i , no one will contribute anything to the group account as in the case with standard preferences. However, for players with positive reciprocity, $r_i > 0$, it is possible to have an equilibrium with positive contributions where $c_i = \frac{r_c(r_i)}{n-1} \sum c_{-i}^{ei} = \frac{r_c(r_i)}{n-1} \sum c_{-i} > 0$. That is, a player with high r_i is more likely to condition his/her contribution on beliefs about others' average contributions, i.e., to be a conditional cooperator. In this case, own contribution is a function of (1) own reciprocity and (2) beliefs about others' average contributions, which are in turn related to their beliefs and reciprocity (c_i is a function of r_i , and hence, c_{-i}^{ei} is a function of r_{-i}^{ei}). Although beliefs about others' reciprocity play a similar role in both games, a difference between the VCM and the trust game is that the latter requires only one member of each pair to form a belief about the preferences of the other. Neither interaction partner needs to form a belief about the other's belief. In the VCM, in contrast, a lengthy and potentially indefinite chain of beliefs about beliefs may be brought to bear on agents' decisions as each of five symmetrically situated members tries to select a contribution c_i (see again (6)).²⁰

¹⁹ In equilibrium, $r_j^{ei} = r_j$ meaning that i 's beliefs about j 's reciprocity is consistent with j 's actual reciprocity.

²⁰ While in principle there may be no clear end to the number of levels of beliefs about beliefs that may be pertinent, we assume that given limited time, subjects translate their rough impressions about the distribution of r in the population into an estimate of the distribution of $[\sum c_{-i}/(n-1)]$.

The simple predictions above have several implications for our experimental design. First, reciprocity and beliefs about reciprocity play as essential a role in contribution decisions in the VCM as in trusting and trustworthiness decisions in the trust game. Although a plethora of equilibria exist, how subjects form an estimate of what other group members will contribute to the group account is essential to selecting one.

Relatedly, despite the fact that consistency between beliefs and reciprocity is required in equilibrium, having such consistent beliefs may not be trivial in reality due to strategic uncertainty. In particular, given the assumption that most people believe that there is substantial variation of reciprocity within a population, any information that can be instrumentally used to affect one's belief about others' reciprocity may change his/her behavior. For our experimental design, in four of their five VCM interaction environments, subjects have potentially suggestive information about the relative trust or trustworthiness of those they are grouped with, and that information may lead them to form estimates that differ by environment. We note that subjects' belief about others' reciprocity are also likely to be population and perhaps context specific, hence the importance, when assessing results, of taking into account our subject pool, experimental protocol, etc.²¹ Controlling for such concerns in the lab, we expect that the information about other group members' trustworthiness in the trust game can help each group member to form an estimate of others' reciprocity in the VCM because both second mover returning in the trust game and contributing in the VCM are positive functions of own x . This implies that knowing about other members' trustworthiness in the trust game will affect subjects' first-order beliefs about other members' contributions in the VCM, which in turn influences contributing behavior in the VCM.

²¹ For example, as noted, the undergraduate student body at Brown University is ethnically and racially diverse, with a fairly high share of international students from many countries. There is a total of roughly 6,500 undergraduates, and participants not only vary from students majoring in music or literature to physics, engineering, economics or sociology, but also include many for whom this may be a first lab experiment experience and some with one, two or more past experiments (which are unlikely to have included trust games). We offer these details not so much as a guide to predicting exact behaviors, but to be clear that beliefs about others' reciprocity are likely to vary both within our subject pool and relative to other groups of subjects, which is precisely the reason that varying trust or social capital are considered to potentially play explanatory roles for differences in country economic performance.

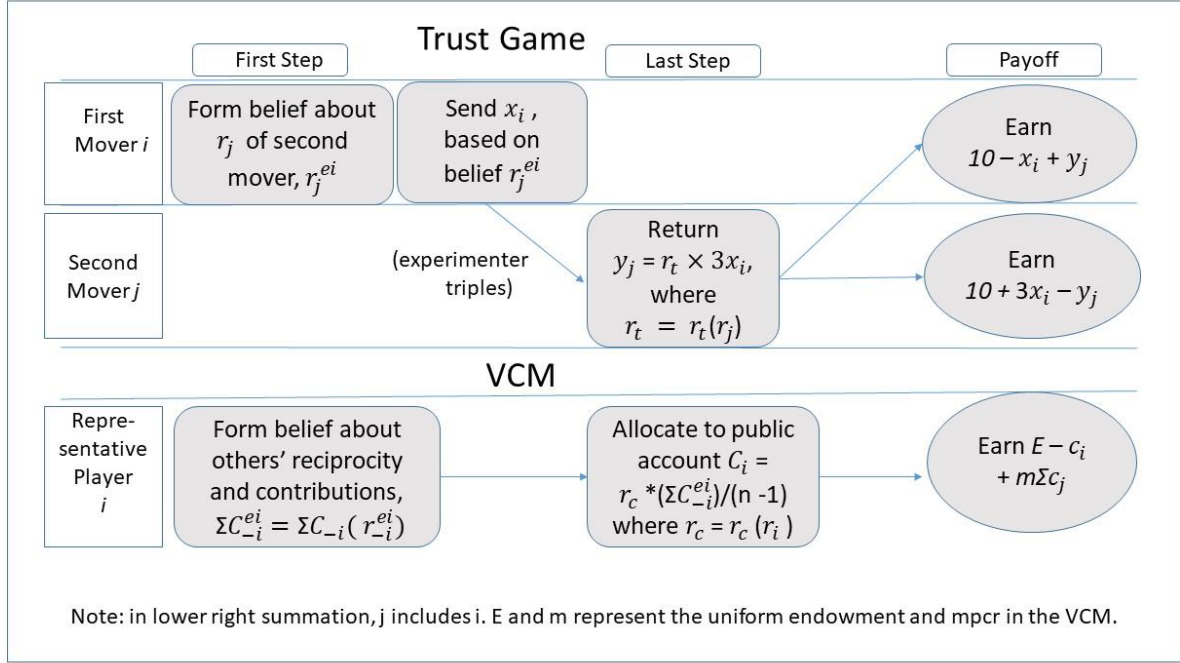


Figure 1. Diagrammatic representation of trust game and VCM.

The analysis so far leads to predictions about behavior in our experiment that we will make concrete in a set of hypotheses in the next sub-section. Figure 1 summarizes schematically how the VCM collapses the two trust game roles into one, and the roles of underlying reciprocal taste r and beliefs about others' r 's in determining behaviors. The top two rows represent the roles of the trust game's first and second movers, respectively, whereas a single row—the third—represents the role of each of the symmetric players in the VCM. The first step in both the trust game and the VCM is formation of a belief about other players, while the last step in both is a costly decision, specifically a returning decision in the trust game and a contributing decision in the VCM. Both of these decisions are determined in large part by the relevant decision-maker(s)' r , with trust game return amount y_j and VCM contribution C_i both being higher the higher is own r , *ceteris paribus*. Important differences between the two games are (a) the roles of belief formation and of reciprocating are specialized to the first and second movers, respectively, in the trust game, whereas both roles are assigned to every player (who is thus an unspecialized actor) in the VCM, and (b) the relevant belief, formed in the first step, is manifested in a distinct, costly move by the first mover, in the trust game—namely sending an amount X —whereas the corresponding belief in the VCM is not indicated by anything separable from the last stage action, although it may be captured in information experimenters elicit regarding beliefs, if an incentive-compatible elicitation succeeds.

Finally, in order to form predictions of experimental behavior and also when applying the framework to other environments with incomplete information, we add an assumption of introspection bias (Butler et al., 2016; Sapienza et al., 2013): an individual's own value of r is

likely to have a significant effect on their belief regarding the counterpart's r (r_j^{ei} is a positive function of r_i), and more so the less specific information about the counterpart is available. In our initial trust games, especially, subjects, being inexperienced in its play and having no feedback about others' actions, will tend to guess what a counterpart will return partly by introspection, so anticipated return varies positively with own r , causing chosen sending to do so as well. Although decisions are made in less of an informational vacuum in our VCM environments, subjects have as yet learned nothing about the range of first and second mover decisions in the trust game, hence their beliefs regarding how relatively high and relatively low trustworthiness and trusting will have been manifested, and how they will translate into VCM contributions, are still likely to be affected by own r , even though a pattern of expecting average r_j^{ei} to be higher in the high than in the low trustworthiness VCM environment, and somewhere in between for the randomly chosen group, is likely. We likewise expect subjects to "project" introspection bias onto others, and therefore to assume that an individual who returns a larger share in the trust game is likely to also send more as a first mover. However, we make no prediction about the degree of correlation subjects will assume on this last matter. It is consistent with our theoretical framework that subjects could view indications of trustworthiness as more reliable predictors of cooperation than are corresponding indications of trust, since the trusting decision is a further step removed from own r , according to our framework. What they tell us about subjects' assessments of how indicative trustworthiness is of trusting is one of the interesting features of our data, given its theoretical unpredictability.

4.2. Hypotheses

We next lay out our hypotheses regarding behaviors in our experiment using the framework above plus three auxiliary assumptions about beliefs. First, we assume that the level of reciprocity varies within any given population. Second, we assume that most experiment participants believe that the level of reciprocity varies within our population of subjects. Third, we assume that subject beliefs regarding the average level of reciprocity in the population vary in a manner correlated with own level of reciprocity, so that for any two individuals j and k such that $r_j > r_k$, j 's belief about the average of r is higher than is k 's when provided with the same incomplete information about that population, again reflecting introspection bias. The first set of hypotheses tests our assumption of introspection bias and the tight connection between behaviors in the trust game and the VCM.

Hypothesis 1 (trust and trustworthiness): *The amount sent as first mover in the trust game is a positive function of the amount returned as second mover.*

Hypothesis 2 (trustworthiness and first order beliefs in the VCM): *Subjects who as trust game second movers return relatively large proportions have higher first-order beliefs in the VCM.*

Hypothesis 3 (trust and first order beliefs in VCM): *Subjects who as first mover send a relatively large amount in the trust game have higher first-order beliefs about others' contributions in the VCM, ceteris paribus.*

H1 follows from the fact that subject j chooses larger X the larger does she estimate the r_i of her counterpart to be, and that the latter estimate is positively correlated with j 's own r . H2 has its most direct application to the environment in which group members are chosen randomly. In that case, an individual has no relevant information about which others are in the group, and thus can only use her belief about the general distribution of r in the population to form a belief about others' contributions. In consequence, H3 follows from H1 and H2.

The next hypothesis deals with the main prediction of our framework regarding the information about others' relative trust/trustworthiness, first-order beliefs and contributions in the VCM.

Hypothesis 4 (first-order beliefs and contributions in the VCM): *Subjects have higher (lower) first-order beliefs about others' average VCM contribution when in a high (low) trust or high (low) trustworthiness environment than when in a randomly formed group. Because many subjects have $r_c > 0$, subjects on average contribute more (less) to the group account when in high (low) environments than when in a random environment.*

H4 follows from the framework laid out in conjunction with Figure 1 and our assumption in this section that most subjects believe that r varies within the population. The belief that r varies implies that a group in which others were in the session's top (bottom) ranks for trustworthiness (trust) is expected to include individuals of higher (lower) r , with average r being of intermediate value in a randomly chosen group. Since higher (lower) r also predicts higher VCM contribution, according to our framework, the estimate of $[\sum C_i / (n - 1)]$ will vary as indicated.

H4 supports the idea that on average, trust and trustworthiness in the trust game affect contributing behavior in the VCM through a channel of forming different first-order beliefs about others' contributions across different environments. We note, however, that the effect of belief on own contribution is by no means uniform. Although subjects having high r are expected to positively adjust their VCM contribution according to their belief about other group members, which varies by environment, subjects for whom $r_c = 0$ are expected not to contribute regardless of environment. The next hypothesis deepens our understanding of the relationship between reciprocity and conditional cooperation by looking at the individual level analysis of whether reciprocity in the trust game can predict the behavioral types in the VCM.

Hypothesis 5 (trustworthiness and conditional cooperation) *Subjects who returned a relatively high proportion as trust game second movers are more likely to be conditional cooperators in the VCM than are subjects who returned nothing, who in turn are more likely to be free riders in the VCM.*

A different possible link between behavior in the trust game and VCM contribution is guilt aversion, which acts through second- rather than first-order beliefs about contribution. Following the psychological game framework of Dufwenberg *et al.* (2011), individuals' decisions can be influenced by their beliefs regarding others' choices, which are in turn indicators of those others' intentions of kindness or unkindness towards them. At the same time, individuals' decisions can also be influenced by their beliefs regarding others' beliefs, because they would incur guilt by letting those others down.²²

Hypothesis 6 (second-order beliefs, guilt aversion and VCM contributions): *Subjects have higher (lower) second-order beliefs about others' average VCM contribution when in a high (low) trust or high (low) trustworthiness environment than when in a randomly formed group. Consequently, guilt aversion leads subjects to contribute more (less) to the group account when in high (low) environments than when in a random environment.*

Regarding H4 and H6, it is an empirical question of which order of beliefs is a more prominent driving force for reciprocity in the trust game to determine contributing behavior in the VCM. We will investigate this question below.

5. Results and Analyses

A total of six experimental sessions were conducted between November and December 2015 at Brown University Social Science Experimental Laboratory (BUSSEL) using the experimental software z-tree (Fischbacher, 2007). For each session, 20 subjects from the university's diverse undergraduate student body (including numerous international students and a wide variety of majors) were recruited by email invitations to the BUSSEL registrants list. 59% of subjects were female, slightly above the female share of the overall student body. Each session lasted around 1.5 hours, and the average earnings were \$18.20, including a \$5 show-up fee.

5.1. Descriptive behavior in Trust Game & Voluntary Contribution Mechanism

We begin by presenting a description of behavior in the trust game and in the VCM. On average, our subjects sent 19.6 tokens (39.2% of their endowments) as senders in the trust game, with a standard deviation of 15.4 tokens. When in the role of recipients in the condition of having been sent all 50 tokens, subjects on average returned 48.7 tokens, 32.5% of the tripled amount received, with a standard deviation of 40.9 tokens. The average returning percentages range from 30.0 – 32.5% for all possible 11 amounts that can be sent. These observations in our trust game are within the range that has been observed in previous experiments, albeit a little below average.²³ Our first

²² Appendix 1 presents a theoretical framework of following Dufwenberg *et al.* (2011) more formally.

²³ See Chaudhuri (2008) and Fehr (2009) for reviews, and see Johnson and Mislin (2011) for a meta-analysis of the trust game experiments. Our subjects' sent amount is 0.85 standard deviations below the average sent amount in 65 studies included in Johnson and Mislin (2011)'s analysis, with a not insubstantial fraction of those studies reporting lower average sent amounts than ours. Our subjects' average returned amount is also on the low side relative to the average of 36.5% in the studies considered by Johnson and Mislin (2011), but a smaller 0.43 standard deviations below the average, and similar to behaviors in the original study by Berg *et al.* (1995), in which, as with our data, it

empirical question is whether as assumed in Hypothesis 1, there exist positive correlations between the amount sent as first mover and the amount returned as second mover. The Spearman's correlation between subjects' first mover sending (as percentage of 50 token endowment) and average return percentage is 0.5629 (p -value < 0.01), a strongly positive association between own trusting and trustworthiness behavior that confirms H1.²⁴

Turning to VCM decisions, Figure 2 shows that in Phase 2, subjects contributed an average of 8.60 of their twenty tokens (43% of their endowments) when in the randomly grouped environment, with a standard deviation of 6.96. In the high trust and low trust environments, the average contributions are 11.48 and 4.9, respectively, with standard deviations of 7.25 and 5.75. In high trustworthiness and low trustworthiness environments, the contributions average 11.22 and 4.60, with standard deviations of 7.07 and 5.57, respectively. Compared to the contributions in the random environment, subjects contributed substantially more in both high trust and trustworthiness environments and less in low trust and trustworthiness environments, respectively, with all the differences significant according to Mann-Whitney tests (p -values < 0.01). The roughly 7-token difference in contribution constitutes an approximately 140% increase in cooperation when subjects move from ostensibly low trust/trustworthiness to a high trust/trustworthiness environment. This indicates that the differing information about trust and trustworthiness rankings has large effects on contributing behavior in the VCM, consistent with H4.²⁵

was not ex post profitable to have sent money (our 32.5% average return share is slightly below the 33.3% required to “make the sender whole”).

²⁴ The reported correlation is calculated using individuals' percent returned averaged over all potential received amounts, but the qualitative finding holds for amount returned at any of those amounts taken individually, with Spearman's correlations between the percentage of amount sent and the percentage of amount returned in each contingent case ranging from 0.47 to 0.59, all associated with p -values less than 0.01.

²⁵ We emphasize here that when subjects made their Phase 2 decisions, their information about trust game decisions was limited to their own choices, only, and thus they had no basis for rejecting the possibility that the first and second mover rankings were entirely randomly assigned to subjects who in fact had no variation in their behaviors. Although the considerable variation in choice shown in Figure 2 strongly suggests beliefs that the trust game behavior rankings sort individuals who vary considerably in type, we lacked direct evidence on this until we decided to ask participants in our final two sessions to estimate the sending and returning behaviors of high and low ranked individuals. We requested the estimates immediately after Phase 2, without prior notice of this request in the experiment instructions, but with incentivization via payment of \$1 for a guess that is accurate within one token. Subjects' average estimate for “what's the average of the 5 lowest sending in Phase 1” and “what's the average of the 5 highest sending in Phase 1” (always within their own session of 20 participants) are 8.12 and 35.31, and the corresponding average estimates for returning when sent all 50 tokens are 7.5 and 85, respectively. This confirms that subjects inferred sharply varying Phase 1 behaviors, although their inferences were entirely “home grown.”

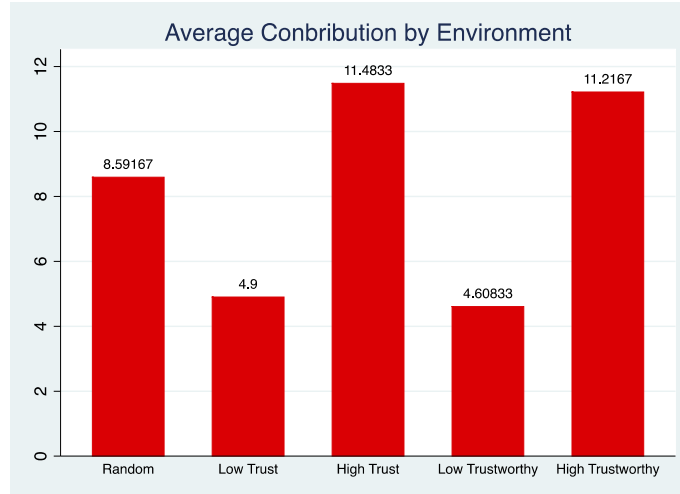


Figure 2. Average Contribution in the VCM in the 2nd Phase

Result 1: Subjects contribute significantly more in high “trust”/“trustworthiness” environments than in low “trust”/“trustworthiness” environments, although they are provided with information solely about members’ relative rankings in the trust game.

In contrast to the significant differences in contributions between high and low (relative) trust and trustworthiness environments, the contribution difference between high trust and high trustworthiness environments (0.26 tokens) is negligible and statistically insignificant; the difference between contributions in low trust and low trustworthiness environments is likewise small (0.29 tokens) and insignificant. These results suggest that subjects treat the relative trusting and the relative trustworthiness of others as interchangeable indicators of underlying type, at least insofar as cooperation is concerned.²⁶ While consistent with our own assumption of introspection bias as well as with our assumption that subjects share common knowledge of the relationships in our modeling framework, the finding is not a trivial one: documenting that people themselves believe that trusting and trustworthiness in others are signs of a common, underlying trait, seems at least as important for research on trust as is observing that trustworthiness and trusting actions are in fact correlated, across individuals.²⁷ To summarize:

²⁶ This is also corroborated by responses to an exit survey question in which we asked subjects to indicate their beliefs as to whether Phase 1 first mover and second mover choices were highly positively correlated, highly negatively correlated, or uncorrelated: out of the 80 subjects that were asked this question (which was not incentivized and was added to the exit survey from which it had been missing in the initial two sessions) over 64% indicated the belief that the behaviors are highly positively correlated (specifically, they rated the degree of correlation based on their own beliefs as either 3 or 4 on a scale from 0 = perfectly negatively correlated to 4 = perfectly positive correlated, with 1 and 3 representing mild negative and positive correlation and 2 uncorrelated. We asked for evaluations according to this informal scale since our subject pool was diverse and we did not want to presume familiarity with formal statistical measures.)

²⁷ One can plausibly imagine, for example, a class of selfish and rational individuals who believe others to be less selfishly rational than themselves and who, in the interest of own payoff maximization, thus exploit reciprocating

Result 2: When making decisions in our VCM games, subjects treat relative trust and relative trustworthiness of other group members interchangeably.

5.2. Beliefs and contributions in the VCM

Our theoretical framework predicts that the main force driving differences in contribution across environments is the extent to which different first-order beliefs about other group members' average contributions are formed in each environment in the VCM. The left panel of Figure 3 depicts average first-order beliefs regarding other group members' average contributions in each environment. Average beliefs track average contribution levels closely in all environments. In high trust and high trustworthiness environments, subjects hold significantly higher beliefs about others' contributions than in the random assignment environment, and the corresponding beliefs in low trust and low trustworthiness environments are significantly lower as predicted by H4 (Mann-Whitney test, p -values < 0.001). To summarize:

Result 3: In environments where others' average first and second mover ranks for sending and returning in the trust game differ appreciably, subjects' beliefs about those others' VCM contributions differ sharply.

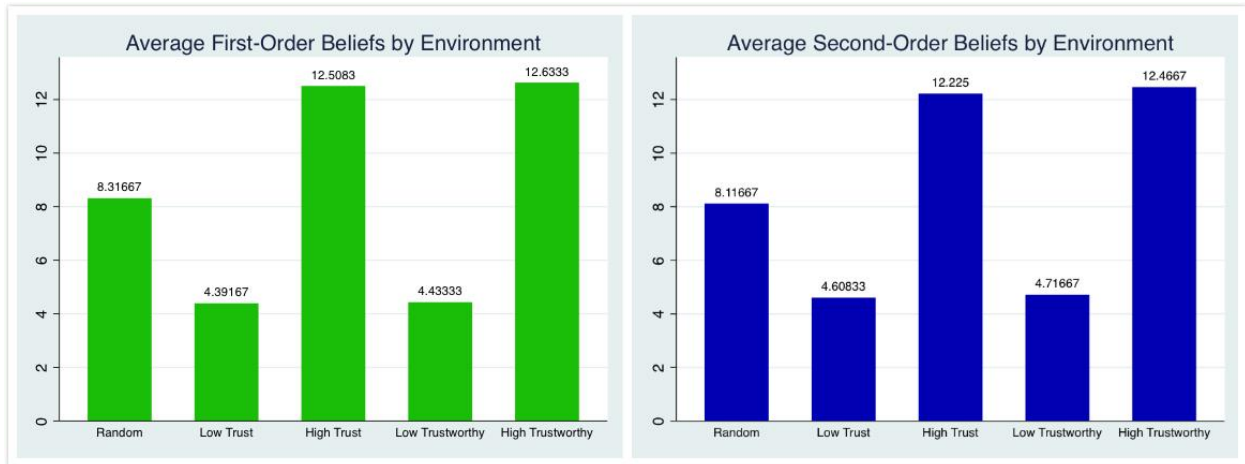


Figure 3. Average beliefs about others' contributions in the VCM, by environment

The right panel of Figure 3 also shows that second-order beliefs well resonate with contributions across environments. Since both the first order beliefs (shown in Fig. 3's left panel) and the second order beliefs (shown in its right panel) are correlated with the contributions shown

second movers by sending their entire endowment when in first mover role, and also return nothing when themselves in second mover role. These individuals would never contribute in a one shot VCM, so their low rank with regard to second mover returning could be a good predictor of their VCM contribution, but their high rank as first mover would be uninformative. As shown later, our subjects' choices suggest instead that scenarios like this play little or no role in their belief formation process.

in Figure 2, it is an empirical question which order of beliefs plays a more important role in determining contributions across different environments. To answer this question, we use regressions in which the dependent variable is individual i 's contribution and the independent variables include both i 's first-order belief (about other members' average contribution) and i 's second-order belief (about other members' average belief regarding the contributions of group members apart from themselves). Table 2 shows the results of a separate linear regression with robust standard errors for each of the five environments, and one in which the contributing behavior of all environments are pooled.²⁸ In each regression, first-order beliefs obtain highly significant positive coefficients, supporting H4, whereas the coefficients on second-order beliefs are small, negative, and quite insignificant, lending little support to H6. This provides a first indication that the association between first-order beliefs and contributions may be stronger than that between second-order beliefs and contributions.²⁹

Table 2. Contribution and Beliefs in the VCM

Dependent Variable:	(1)	(2)	(3)	(4)	(5)	(6)
Contributions	Random	Low Trust	High Trust	Low Trustworthiness	High Trustworthiness	Pooled
First-Order Beliefs	1.081*** (0.204)	1.027*** (0.144)	0.859*** (0.113)	1.156*** (0.0980)	1.158*** (0.122)	0.982*** (0.073)
Second-Order Beliefs	-0.0642 (0.280)	-0.000773 (0.0906)	0.176 (0.142)	-0.0513 (0.107)	-0.117 (0.196)	-0.046 (0.082)
Constant	0.119 (0.662)	0.394 (0.745)	-1.407 (0.730)	-0.274 (0.371)	-1.951 (1.192)	0.241 (0.328)
Observations	120	120	120	120	120	600
R-squared	0.522	0.477	0.448	0.591	0.529	0.579

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

To further check whether the insignificance of second-order beliefs signals genuine lack of impact once first-order beliefs are controlled for, we also use a 2-step differencing approach wherein we first regress second-order beliefs on first-order beliefs, and we save the residuals

²⁸ In this as well as in future regression tables, unless specified, demographic control variables (race/ethnicity, gender, year at university, and major) are insignificant, and including them does not qualitatively affect our results, so we report estimates of specifications that omit them.

²⁹ For completeness, we note that if we include only first-order beliefs and a constant, or only second-order beliefs and a constant, in simpler regressions, each of the two belief variables obtains similarly significant and positive coefficients. In other words, each belief by itself is a strong positive predictor of contribution. Only by including both simultaneously do we obtain evidence that one has stronger predictive power than the other.

because they capture the part of second-order beliefs that is orthogonal to first-order beliefs. Next, in the second step, we regress contribution on first-order beliefs as well as the residuals. If the coefficient of the residuals is statistically significant, then it can be concluded that second-order beliefs separately impact contribution decisions in the VCM. However, the results are quantitatively and qualitatively similar to Table 2: the only significant coefficient is that of first-order beliefs. The 2-step regression outputs for the pooled sample are shown in Table 3.³⁰

Table 3. 2-step regression for beliefs and contributions in the VCM

Dependent Variable	1 st stage Second Order Belief	2 nd stage Contribution
First-Order Belief	0.809*** (0.0249)	0.945*** (0.0334)
Residual from the first stage		-0.0461 (0.0827)
Constant	1.588*** (0.203)	0.168 (0.305)
Observations	600	600
R-squared	0.736	0.580

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Result 4: Subjects' VCM contributions in each environment are a significant positive function of their first-order beliefs, but are not affected by their second-order beliefs once first-order beliefs are controlled for. Contributions thus appear to respond to how much others are expected to contribute rather than to how much others are assumed to believe others including oneself are contributing. This result provides strong support for the role of first-order beliefs and reciprocity in determining contributions in the VCM, consistent with H4, while our data do not appear to support the idea of guilt aversion and the associated H6).

As a robustness check for the results above, a more direct approach is adopted to demonstrate that between-environment differences in first-order beliefs can explain the corresponding differences in contributions. In Table 4, we present estimates of fixed effect models in which we regress the contribution difference, for each subject i , on the difference of i 's first-order beliefs, in high vs. low (trust/trustworthiness) environments. Column (1) shows regressions using differences in first-order beliefs and differences in contributions between the high and the low trust environments, while column (2) shows the corresponding regressions for differences in high vs. low trustworthiness environments. These estimates show that differences in first-order beliefs significantly predict differences in contributions in both trust and trustworthiness environments. The greater the difference in others' contributions i estimates there to be between the high and the low (trust/trustworthiness) environment, the greater the difference in i 's actual contribution in the high vs. in the low corresponding environment. In column (3), we pool the data

³⁰ As a consequence, our analysis focuses mainly on first-order beliefs in the remainder of this paper.

over high and low environments, making use of the apparent similarity of the effects of relative trust and of relative trustworthiness on subject beliefs and behaviors. We obtain findings paralleling those in columns (1) and (2).

Table 4. Difference in Contributions Predicted by Difference in Beliefs

Dependent Variable:	(1)	(2)	(3)
Difference in Contribution	Trust	Trustworthiness	Pooled
Difference in First-Order Beliefs	0.606*** (0.128)	0.717*** (0.152)	0.657*** (-0.069)
Constant	1.663 (0.936)	0.728 (0.695)	1.232* (0.670)
Fixed Effect	Yes	Yes	Yes
Observations	120	120	240
R-squared	0.239	0.318	0.274

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Dependent variable is difference in VCM contribution of subject i when in high trust (trustworthiness) vs. low trust (trustworthiness) environment, while independent variable is difference in first order belief of i regarding other group members' contributions in the high trust (trustworthiness) vs. in the low trust (trustworthiness) environment.

Result 5: The “high trust” and “high trustworthiness” environments promote cooperation more than do their “low trust” and “low trustworthiness” counterparts by inducing individuals to hold higher first-order beliefs about the contributions of other group members.

It is important to note that the results in Table 4 can be interpreted as the marginal relationship between first-order beliefs and contributions in the VCM. The coefficients indicate that a unit increase in beliefs is associated with an increase in contributions of less than one unit. Such an average increase would be consistent with the presence of more than one subject type, for instance perfect conditional cooperators who raise their contribution by a full unit when anticipating that others are doing so, imperfect conditional cooperators who raise their contribution in the same circumstance by only part of a unit, free riders who contribute nothing regardless of their belief about others.³¹ However, the size and high significance of the coefficients suggests that

³¹ These estimated coefficients are also consistent with the findings in the literature of “conditional cooperation” in which conditional cooperators are imperfect in the sense that the degree of reciprocation does not fully match others’ “kind behavior” (see Fischbacher *et al.*, 2001; Fischbacher and Gächter, 2010).

conditional cooperators of some kind are likely to be quite numerous. We investigate presence of conditional cooperation more directly in subsection 5.4.

5.3. Trust, trustworthiness, and first-order beliefs

We have shown so far that first-order beliefs about other group members' contributions are the main driving force in explaining the different levels of contributions across environments in the VCM. We next investigate whether, as Hypotheses 2 and 3 state, subject i 's sending and returning behaviors in the trust game are systematically related (by virtue of introspection bias) to i 's beliefs about others, in the VCM. Table 5 displays the results of a set of regressions in which for each environment, the dependent variable is i 's first-order belief about others' contribution in the VCM, and the explanatory variables are i 's decisions as trustor and trustee in the trust game.³²

Table 5. Reciprocity in the Trust game and First-order beliefs in the VCM

Dependent Variable:	(1)	(2)	(3)	(4)	(5)	(6)
First-order beliefs	Random	Low Trust	High Trust	Low Trustworthiness	High Trustworthiness	Pooled and Differenced
Sending Percentage	1.073 (1.923)	2.364* (0.948)	0.677 (1.698)	2.124** (0.657)	1.236 (2.219)	1.495** (0.682)
Avg. Returning Percentage	2.447* (1.062)	3.134** (0.779)	4.576* (1.988)	2.533*** (0.614)	4.873** (1.691)	3.513*** (0.916)
Constant	7.133*** (0.679)	2.488*** (0.446)	10.82*** (0.742)	2.811*** (0.340)	10.63*** (0.895)	-1.681*** (0.324)
Observations	120	120	120	120	120	600
R-squared	0.069	0.116	0.063	0.086	0.091	0.069
p-value of Wald Test	0.0107**	0.0091***	0.0348**	0.0014***	0.0085***	0.0006***

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The reported Wald Test is for the hypothesis that the coefficients on the sending percentage and average returning percentage variables are jointly statistically significant.

³² Note that we control for i 's trust game decisions in both roles, simultaneously. This poses the danger that the significance of the individual coefficients will be underestimated due to the correlation between the two decisions, but has the advantage that we may be able to determine which of the two, trusting or being trustworthy, (if either) has a stronger impact on beliefs.

First, we find that in keeping with H2, trustworthiness (avg. returning %) is positively and at least marginally significantly correlated with first-order beliefs in all environments. In column (1) – (5), those who return more in the trust game tend to have more optimistic beliefs about others’ contributions in all environments of the VCM than those who return less. Sending in the trust game is also positively correlated with first-order beliefs, as H3 predicts, although its associations are less significant than those between returning behavior and first-order beliefs—a finding recapitulated (except in the Low environments) in regressions that include sending only.³³ For the robustness of this result, we also look at the pooled data across all environments in column 6. To control for differences in environments, we take the dependent variable as the difference between subject i ’s average first-order belief and the average first-order belief of all participants, in the given environment, and find the consistent result. Overall, the results are strongly consistent with the idea that one channel through which trust and especially trustworthiness affect contributions in the VCM is by influencing beliefs about what others will contribute.

Result 6: Subjects who send and return more (less) in the trust game are significantly more (less) optimistic about how other subjects will behave in the cooperation dilemma that is the VCM. The relationship between trustworthiness and first-order beliefs is stronger than that between trust and first-order beliefs.

5.4. Trust, Trustworthiness, and Conditional Cooperation

So far we find a strong average tendency of conditional cooperation, insofar as higher beliefs about what others will contribute are associated with higher contributions in the VCM. However, past research suggests that not all individuals are conditional cooperators, and our predictive framework has allowed that there can be subjects for whom $r_c = 0$, who would be expected not to contribute regardless of environment. We now investigate whether reciprocity in the trust game is correlated with behavioral types in the VCM, as predicted by H5.

Past classifications of conditional cooperation type (Fischbacher *et al.*, 2001, Thöni and Volk, 2018) have been based on subjects’ completions of tables in which they are asked to choose

³³ We estimated variants of the regressions shown in Table 5 but including only sending or only returning, in order to assess the correlation of each with beliefs free of its collinearity with the other (the table is not shown, to save space). In these estimates, returning in the trust game is highly significantly correlated with first-order beliefs in all environments (p-values < 0.05), but positive correlations between sending and first-order beliefs are significant only in the low environments (p-values < 0.05). For the Random and High Trust environments, where only the coefficient on trustworthiness (returning percentage) is statistically significant and only at the 10% level, in Table 5 itself, we also implement Wald tests and find that the coefficients on sending and returning together are jointly significant at the 5% level. Wald tests for the other regressions show joint significance of 1% or better. The less robust correlations of sending than of returning, with respect to VCM beliefs, suggest that there are differences in the behavioral mechanisms behind sending and returning behaviors. Recall that our theoretical framework assumes that trustworthiness is a direct function of underlying reciprocity (r), with trust linked to it only insofar as the actor takes own trustworthiness as a proxy for that of others (introspection bias). Factors besides own r that can affect amount sent as first mover could include experiential inferences from outside the lab regarding others’ trustworthiness (which may be perceived as greater or less than one’s own), altruism (valuing the others’ payoff), risk aversion, and betrayal aversion. Only the first of these seems likely to also affect VCM beliefs.

an amount they will contribute given each possible average amount others might contribute. These choices are incentivized by the fact that in one game round, a randomly selected group member's payoff-determining contribution will be the applicable conditioned one. Our experimental design provides a different kind of data for classifying conditional contributing tendencies: from the pairings of elicited first order belief and own contribution choice in our five VCM environments. This has the potential advantage that decisions are made without the potentially structuring frame of a table, and with beliefs and contributions inputted on separate screens, but the disadvantage that contribution decisions are taken for a maximum of five amounts others might contribute instead of the typical 11 or 21, and with the set of amounts covered varying from subject to subject and potentially covering a smaller range. These limitations lead us to slightly modify the classification scheme used in Fischbacher *et al.* (2001). We start by identifying 5 categories: (1) free rider, (2) weak conditional cooperator, (3) strong conditional cooperator, (4) altruist, and (5) other.³⁴ 27 subjects cannot be assigned to one of the first four categories and are thus classified in category (5), which we exclude from the analysis that follows due to its high internal diversity and lack of compelling placement relative to categories (1) to (4). This leaves us with 93 out of 120 subject observations for our initial exercises. We also consider an expanded classification scheme that permits sixteen of the 27 “other” types to be assigned to one or another of seven categories, which add to the above-listed five types a type (1'), “broad free riders” (inserted between types (1) and (2) above), and type (1''), “hump-shaped contributors” (also before type (2) above, but after type (1')).^{35, 36} Both variants of our classification exercise show our modal subject to be conditionally cooperative, as in other studies.³⁷

Having classified most subjects in terms of a reciprocity type in the VCM, we can now test H5, which captures a core assumption of our analytical framework, namely that individuals can be characterized by a reciprocity parameter that predicts both trust game returning decisions and cooperation conditional on first-order beliefs in the VCM. With each set of categorical variables, we estimate ordered multinomial logit regressions as shown in Table 6 to check whether behavioral

³⁴ We classify subjects by graphing their contribution against first-order beliefs in each environment. If the individual contributes nothing in all environments, regardless of his/her beliefs, he/she is classified as a free rider. If the individual always contributes at least as much as his/her belief about the average amount that others are contributing in the group, he/she is a strong conditional cooperator. If the individual contributes at least 4 out of 5 times when others are expected to contribute a positive amount on average, but doesn't always contribute as much as the belief and does not show a hump shaped pattern (see below), the individual is a weak conditional cooperator. If the individual contributes all 20 tokens in all environments, we call this individual an altruist. All other subjects are classified as “other”.

³⁵ Broad free riders are subjects who contribute 0 or 1 in most conditions and never contribute more than 5. Hump-shaped contributors are a type initially defined by Fischbacher *et al.* (2001) whose contribution initially increases as others' average contribution rises, then beyond some level of contribution by others begins to fall (a possible interpretation being that this type of individual is willing to help get cooperation off the ground, but shifts towards free riding when others seem cooperative enough to be in less need of this ‘support’).

³⁶ Numbers of subjects by classification type are: (1) Free rider: 18, (1') Broad free rider: 3, (1'') Hump-shaped: 7, (2) Weak conditional cooperator: 39, (3) Strong conditional cooperator: 39, (4) Altruist: 6, and (5) Other: 8. For multinomial logit regression analysis, we renumber the sequence 1, 1', 1'', 2, ..., 5 as 1, 2, 3, ..., 7.

³⁷ In Fischbacher *et al.* (2001), 50% of their subjects are conditional cooperator, 30% free rider, 14% hump shaped, and 6% others. Compared to Fischbacher *et al.* (2001), we have slightly more cooperators and less free riders in our subjects: 65% are conditional cooperator (both strong and weak), 17.5% free rider (both strict and broad), 6% hump shaped, 5% altruist, and 6.5% others, although correspondences are inexact due to slight differences in assignment criteria. See Thöni and Volk (2018) for a meta-analysis of 17 replication studies based on the Fischbacher *et al.* design.

disposition in the VCM can be predicted by 1st or 2nd mover choices in the trust game. The results are shown in Table 6, with columns (1) – (3) using the 4 type classification that covers 93 subjects, and columns (4) – (6) using the 6 type classification covering 109 subjects.

Table 6. Trust, trustworthiness, and conditional cooperation.

Dependent Variable : VCM Types, Ordered	4 type specification			6 type specification		
	(1)	(2)	(3)	(4)	(5)	(6)
Sending Percentage in the Trust Game	0.751** (0.373)		0.213 (0.468)	0.700** (0.337)		0.370 (0.395)
Average Returning Percentage in the Trust Game		1.295*** (0.474)	1.134* (0.591)		1.040** (0.423)	0.801 (0.494)
Cut point 1	-0.724*** (0.206)	-0.646*** (0.203)	-0.614*** (0.215)	-0.846*** (0.193)	-0.800*** (0.193)	-0.738*** (0.205)
Cut point 2	0.337* (0.197)	0.450** (0.199)	0.483** (0.212)	-0.761*** (0.190)	-0.714*** (0.191)	-0.650*** (0.203)
Cut point 3	1.835*** (0.260)	1.965*** (0.264)	1.999*** (0.276)	-0.447** (0.184)	-0.398** (0.184)	-0.331* (0.198)
Cut point 4				0.479*** (0.182)	0.533*** (0.183)	0.605*** (0.198)
Cut point 5				1.895*** (0.245)	1.957*** (0.247)	2.031*** (0.261)
Observations	93	93	93	109	109	109
Pseudo R-Squared	0.018	0.0338	0.0348	0.0137	0.0193	0.0221
p-value of Wald test	-	-	0.0148**	-	-	0.0042***

Note: for columns (1) – (3), types are: 1=free rider, 2=weak conditional cooperator, 3=strong conditional cooperator, and 4=Altruist.

For columns (4) – (6), types are: 1=free rider, 2=broad free rider, 3=hump shaped contributor, 4= weak conditional cooperator, 5=strong conditional cooperator, and 6=Altruist.

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The reported Wald Test is of the hypothesis that the coefficients on the sending percentage and average returning percentage variables are jointly statistically significant.

In column (1) – (3), we find that sending and returning behavior in the trust game are strongly correlated with types in the VCM ordered from free rider to strong conditional cooperator and altruist. Roughly speaking, subjects who return more in the trust game are less likely to be free riders and more likely to be strong conditional cooperators and altruists in the VCM. Regarding the marginal effects, a 1% increase in returned amount is associated with a 0.32% and 0.13% increase on the probability of being a strong conditional cooperator and an altruist at the 10% significance level, respectively, whereas the same changes in returning behavior leads to a decrease in the probability of being a free rider and a weak conditional cooperator by 0.05% and 0.04% at the 10% significance level, respectively. The estimates in columns (4) – (6) are similar, with both trusting (column 4) and trustworthiness (column 5) individually predicting reciprocity type in the VCM, and with the two variables jointly significant at the 1% level in column (6), according to the Wald test. Therefore, we end this section with the following conclusion:

Result 7: Subjects' behavior in the trust game is a good predictor of their type in the VCM at the individual level. In particular, those who send and return more in the trust game are more likely to be altruists or conditional cooperators than those who send and return less.

6. Discussion and conclusion

The importance of trust to economic activity is much remarked, but the channels through which trust helps to lift economic output are still imperfectly understood. We suggest that one reason greater trust is associated with better economic outcomes is that greater trust is a response to greater trustworthiness, and that trustworthiness is a reflection of the same inclination to reciprocate others' kind or cooperative behavior as is cooperation in the simultaneous move social dilemmas that characterize partnerships and teams, provision of local public goods, and civic engagement and monitoring of public institutions. To the extent that they prove indicative of broader populations, our data suggest that most people treat signs of trust and trustworthiness as indicators that it is safe to cooperate in a simultaneous move social dilemma in the sense that one's cooperation is most likely to be matched by that of others, these others also likely being conditional cooperators. We tested for presence of such a pathway from indications of trust and trustworthiness to expectations of cooperation and decisions to cooperate by asking subjects to make trust game decisions in both roles, then (without prior announcement) to make one shot public goods contribution decisions in environments differing in terms of the average ranking of fellow group

members with respect to trust (first mover sending) and trustworthiness (second mover return proportion). Ours is the first investigation of how information about trust and trustworthiness is translated into expectations about cooperation, and the first study of whether the two sides of trust and trustworthiness are treated differently when forming such expectations. Along the way, our study also yields novel findings about (i) “home grown priors” concerning the variation of trusting and trustworthiness among a set of anonymous others, and about (ii) the relative importance of guilt aversion and reciprocity in driving voluntary contributions. We also confirm a strong correlation between trust and trustworthiness in given individuals, as have other studies.

As anticipated from past experiments, most subjects chose to contribute positive amounts in the VCM games, and on average they contributed more the more that they expected others in the group in question to contribute, consistent with the modal subject type being a conditional cooperator. We elicited incentivized first and second order beliefs about contributions in VCM groups varying by relative trust and trustworthiness, performing this elicitation after rather than before or simultaneous with contribution decisions so as to avoid contamination of choices by belief elicitation. We found that subjects on average anticipated much higher contributions in groups of higher ranked trustors and trustees than in groups of lower ranked trustors and trustees, with expectations for a random group lying in between. This was the case despite subjects knowing that rankings could be random (ties would be broken randomly) if there was no variation in trust game choices. Interestingly, contribution expectations for higher ranked trustors and higher ranked trustees were indistinguishable, as were contribution expectations for lower ranked trustors and lower ranked trustees. In sum, subjects appear to believe that people vary considerably in how they will behave in both a trust game and a VCM. They believe that a sizeable fraction of the population are quite trusting and trustworthy in the trust game and that these same individuals are quite cooperative in the VCM. They believe another sizeable fraction are much less trusting, trustworthy, and cooperative. The fact that their predictions based on relative second and on relative first mover rank in the trust game are interchangeable suggests that the subjects operate with beliefs similar to that of the theoretical approach we presented, i.e. that both share returned in the trust game and amount contributed in the VCM are positive functions of own level of reciprocity, and that decisions about how much to send as a trust game first mover are based on beliefs about how much a second mover will return, which are strongly influenced by introspection and thus by own degree of reciprocity.

Subjects’ second order beliefs (about what others expected other group members to contribute) were highly correlated with their first order beliefs (their own expectation of what other group members would contribute), and both first and second order beliefs, taken individually, were highly correlated with amount sent in most VCM environments. However, when both sets of beliefs are simultaneously controlled for, we found that first order beliefs significantly predicted contributions, with second order beliefs (or, alternatively, their extracted orthogonal component) not a significant predictor of contributions. This result supports a reciprocity explanation of VCM cooperation over the alternative guilt aversion explanation proposed by Dufwenberg *et al.* (2011). Relatedly, when we classify subjects into types based on VCM contributions in differing

environments with differing first order beliefs, we find that being more (less) trustworthy in the trust game is a significant predictor of being a conditional cooperator (free rider).

Returning to its broader motivation, application of our experimental results to the posited pathway connecting trust to economic growth depends on the validity of extrapolating from variation of reciprocity type *r within* a society (such as the subject pool in our lab) to variation of *r across* societies. Holding the disposition towards cooperation of any given individual constant, that individual would be expected to behave more cooperatively in dilemmas of group effort, social norm adherence, etc., when living in a society in which a large proportion of others are observed to be trustworthy (in part, perhaps, as indicated by their trust towards one another) and less cooperatively when in a society in which the trustworthy are believed to be few. Since cooperativeness in the workplace, adherence to agreements structuring economic transactions, collective vigilance against corruption in public agencies, and other forms of cooperation can play key roles in maximizing the productivity of resources, the line connecting trust with economic efficiency and growth is likely grounded in the disposition of reciprocity. While the potential for such a disposition to be present in human populations may well have deep roots in human evolution, its strength and incidence in given societies depends on cultural, historical, and institutional factors beyond the scope of our paper. What we have demonstrated seems nonetheless an important piece of the overall puzzle: we have shown that people's willingness to bear the cost of cooperation varies significantly with the environment they believe themselves to be operating in, and that the trustworthy and trusting actions of those in the environment are treated as strong cues about the types of others who comprise it.

References

- Akerlof, George A. "Labor contracts as partial gift exchange." *Quarterly Journal of Economics* 97.4 (1982): 543-569.
- Algan, Yann, and Pierre Cahuc. "Inherited trust and growth." *American Economic Review* 100.5 (2010): 2060-92.
- Algan, Yann, and Pierre Cahuc. "Trust, growth, and well-being: New evidence and policy implications." *Handbook of economic growth*. Vol. 2. Elsevier (2014): 49-120.
- Arrow, Kenneth J. "Gifts and exchanges." *Philosophy and Public Affairs* (1972): 343-362.
- Ashraf, Nava, Iris Bohnet, and Nikita Piankov. "Decomposing trust and trustworthiness." *Experimental economics* 9.3 (2006): 193-208.
- Bellemare, Charles, and Sabine Kröger. "On representative social capital." *European Economic Review* 51.1 (2007): 183-202.
- Ben-Ner, Avner and Louis Putterman, "Trusting and Trustworthiness," *Boston University Law Review* 81.3 (2001): 523-551.
- Berg, Joyce, John Dickhaut, and Kevin McCabe. "Trust, reciprocity, and social history." *Games and Economic Behavior* 10.1 (1995): 122-142.
- Blanco, Mariana, Dirk Engelmann, Alexander K. Koch, and Hans-Theo Normann. "Belief elicitation in experiments: is there a hedging problem?" *Experimental Economics* 13.4 (2010): 412-438.
- Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann. "A within-subject analysis of other-regarding preferences." *Games and Economic Behavior* 72.2 (2011): 321-338.
- Bohnet, Iris, and Richard Zeckhauser. "Trust, risk and betrayal." *Journal of Economic Behavior and Organization* 55.4 (2004): 467-484.
- Bolton, Gary E., and Axel Ockenfels. "A theory of equity, reciprocity and competition." *American Economic Review* 90.1 (2000): 166-193.
- Bonin, John P., Derek C. Jones, and Louis Putterman. "Theoretical and empirical studies of producer cooperatives: will ever the twain meet?" *Journal of Economic Literature* 31.3 (1993): 1290-1320.
- Butler, Jeffrey V., Paola Giuliano, and Luigi Guiso. "Values we learn from our parents influence our trust in others with money and business." *LSE Business Review* (2016).

Charness, Gary, and Martin Dufwenberg. "Promises and partnership." *Econometrica* 74.6 (2006): 1579-1601.

Charness, Gary, and Matthew Rabin. "Understanding social preferences with simple tests." *Quarterly Journal of Economics* 117.3 (2002): 817-869.

Chaudhuri, Ananish. *Experiments in economics: playing fair with money*. Routledge (2008).

Chaudhuri, Ananish. "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature." *Experimental Economics* 14.1 (2011): 47-83.

Chen, Josie I., Andrew Forster, and Louis Putterman. "Identity, Trust and Altruism: An Experiment on Preferences and Microfinance Lending," Working Paper 2017-3, Department of Economics, Brown University (2017).

Cox, James C. "How to identify trust and reciprocity." *Games and Economic Behavior* 46.2 (2004): 260-281.

Dariel, Aurelie, and Nikos Nikiforakis. "Cooperators and reciprocators: A within-subject analysis of pro-social behavior." *Economics Letters* 122.2 (2014): 163-166.

Dufwenberg, Martin, and Georg Kirchsteiger. "A theory of sequential reciprocity." *Games and Economic Behavior* 47.2 (2004): 268-298.

Dufwenberg, Martin, Simon Gächter, and Heike Hennig-Schmidt. "The framing of games and the psychology of play." *Games and Economic Behavior* 73.2 (2011): 459-478.

Eckel, Catherine C., and Rick K. Wilson. "Is trust a risky decision?" *Journal of Economic Behavior and Organization* 55.4 (2004): 447-465.

Ellingsen, Tore, Magnus Johannesson, Johanna Mollerstrom, and Sara Munkhammar. "Social framing effects: Preferences or beliefs?" *Games and Economic Behavior* 76.1 (2012): 117-130.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman and Uwe Sunde, "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences," IZA Discussion Paper 9674 (2016).

Falk Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman and Uwe Sunde, "Global Evidence on Economic Preferences," *Quarterly Journal of Economics* 133.4 (2018): 1645-1692.

Falk, Armin, and Urs Fischbacher. "A theory of reciprocity." *Games and economic behavior* 54.2 (2006): 293-315.

Fehr, Ernst. "On the economics and biology of trust." *Journal of the European Economic Association* 7.2-3 (2009): 235-266.

Fehr, Ernst, and Klaus M. Schmidt. "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics* 114.3 (1999): 817-868.

Fehr, Ernst, and Urs Fischbacher. "The nature of human altruism." *Nature* 425.6960 (2003): 785-791.

Fischbacher, Urs. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10.2 (2007): 171-178.

Fischbacher, Urs, and Simon Gächter. "Social preferences, beliefs, and the dynamics of free riding in public goods experiments." *American Economic Review* 100.1 (2010): 541-556.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics Letters* 71.3 (2001): 397-404.

Fukuyama, Francis. *Trust: The social virtues and the creation of prosperity*. No. D10 301 c. 1/c. 2. New York: Free press (1995).

Gächter, Simon, Benedikt Herrmann, and Christian Thöni. "Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence." *Journal of Economic Behavior and Organization* 55.4 (2004): 505-531.

Gächter, Simon, and Christian Thöni. "Social learning and voluntary cooperation among like-minded people." *Journal of the European Economic Association* 3.2-3 (2005): 303-314.

Geanakoplos, John, David Pearce, and Ennio Stacchetti. "Psychological games and sequential rationality." *Games and Economic Behavior* 1.1 (1989): 60-79.

Gintis, Herbert, *Game Theory Evolving: A Problem-Centered Introduction to Evolutionary Game Theory*, 2nd Edition. Princeton: Princeton University Press (2009).

Glaeser, Edward L., David I. Laibson, Jose A. Scheinkman, and Christine L. Soutter. "Measuring trust." *Quarterly Journal of Economics* 115.3 (2000): 811-846.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. "Cultural biases in economic exchange?" *Quarterly Journal of Economics* 124.3 (2009): 1095–1131

Gunnthorsdottir, Anna, Daniel Houser, and Kevin McCabe. "Disposition, history and contributions in public goods experiments." *Journal of Economic Behavior and Organization* 62.2 (2007): 304-315.

Gul, Faruk, and Wolfgang Pesendorfer. "Interdependent preference models as a theory of intentions." *Journal of Economic Theory* 165 (2016): 179-208.

Hoffman, Elizabeth, Kevin A. McCabe, and Vernon L. Smith. "Behavioral foundations of reciprocity: Experimental economics and evolutionary psychology." *Economic Inquiry* 36.3 (1998): 335-352.

Isaac, R. Mark, James M. Walker, and Susan H. Thomas. "Divergent evidence on free riding: An experimental examination of possible explanations." *Public Choice* 43.2 (1984): 113-149.

Johnson, Noel D., and Alexandra A. Mislin. "Trust games: A meta-analysis." *Journal of Economic Psychology* 32.5 (2011): 865-889.

Keser, Claudia, and Frans Van Winden. "Conditional cooperation and voluntary contributions to public goods." *Scandinavian Journal of Economics* 102.1 (2000): 23-39.

Kőszegi, Botond, and Matthew Rabin. "A model of reference-dependent preferences." *Quarterly Journal of Economics* 121.4 (2006): 1133-1165.

Khalmetski, Kiryl. "Testing guilt aversion with an exogenous shift in beliefs." *Games and Economic Behavior* 97 (2016): 110-119.

Knack, Stephen, and Philip Keefer. "Does social capital have an economic payoff? A cross-country investigation." *Quarterly Journal of Economics* 112.4 (1997): 1251-1288.

Kocher, Martin G., Peter Martinsson, Dominik Matzat, and Conny Wollbrant. "The role of beliefs, trust, and risk in contributions to a public good." *Journal of Economic Psychology* 51 (2015): 236-244.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. "Legal determinants of external finance." *Journal of Finance* 52.3 (1997): 1131-1150.

Lazzarini, Sergio G., Regina Madalozzo, Rinaldo Artes, and Jose de Oliveira Siqueira. "Measuring trust: An experiment in Brazil." *Brazilian Journal of Applied Economics* 9.2 (2005): 153-169.

Ledyard, John O. "Public Goods: A Survey of Experimental Research." In John H. Kzagal and Alvin E. Roth (eds): *The Handbook of Experimental Economics*. Princeton: Princeton University Press (1995): 111-194.

Levine, David K. "Modeling altruism and spitefulness in experiments." *Review of Economic Dynamics* 1.3 (1998): 593-622.

Liberman, Varda, Steven M. Samuels, and Lee Ross. "The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves." *Personality and social psychology bulletin* 30.9 (2004): 1175-1185.

Murtin, Fabrice, Lara Fleischer, Vincent Siegerink, Arnstein Aassve, Yann Algan, Romina Boarini, Santiago González, Zsuzsanna Lonti, Gianluca Grimalda, Rafael Hortala Valllve,

Soonhee Kim, David Lee, Louis Putterman and Conal Smith, "Trust and its Determinants: Evidence from the TrustLab Experiment," OECD Statistics Working Papers 2018/02 (2018).

Ones, Umut, and Louis Putterman. "The ecology of collective action: A public goods and sanctions experiment with controlled group formation." *Journal of Economic Behavior and Organization* 62.4 (2007): 495-521.

Ostrom, Elinor. "Beyond markets and states: polycentric governance of complex economic systems." *American Economic Review* 100.3 (2010): 641-72.

Page, Talbot, Louis Putterman, and Bulent Unel. "Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency." *Economic Journal* 115.506 (2005): 1032-1053.

Putterman, Louis, Jean-Robert Tyran, and Kenju Kamei. "Public goods and voting on formal sanction schemes." *Journal of Public Economics* 95.9-10 (2011): 1213-1222.

Rabin, Matthew. "Incorporating fairness into game theory and economics." *American Economic Review* (1993): 1281-1302.

Sapienza, Paola, Anna Toldra-Simats, and Luigi Zingales. "Understanding trust." *Economic Journal* 123.573 (2013): 1313-1332.

Schotter, Andrew, and Isabel Trevino. "Belief elicitation in the laboratory." *Annual Review of Economics* 6.1 (2014): 103-128.

Selten, Reinhard. "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes." Seminar für Mathemat. Wirtschaftsforschung u. Ökonometrie (1965).

Sen, Amartya K. "Isolation, assurance and the social rate of discount." *Quarterly Journal of Economics* (1967): 112-124.

Tabellini, Guido. "Culture and institutions: economic development in the regions of Europe." *Journal of the European Economic association* 8.4 (2010): 677-716.

Thöni, Christian. "Trust and Cooperation: Survey Evidence and Behavioral Experiments." in Paul A. M. Van Lange, Bettina Rockenbach, & Toshio Yamagishi (eds): *Social dilemmas: New perspectives on trust*. New York: Oxford University Press (2017).

Thöni, Christian, Jean-Robert Tyran, and Erik Wengström. "Microfoundations of social capital." *Journal of Public Economics* 96.7 (2012): 635-643.

Thöni, Christian and Stefan Volk, "Conditional Cooperation: Review and Refinement," *Economics Letters* 171 (2018): 37 - 40.

Zak, Paul J., and Stephen Knack. "Trust and growth." *Economic Journal* 111.470 (2001): 295-321.

Zelmer, Jennifer. "Linear public goods experiments: A meta-analysis." *Experimental Economics* 6.3 (2003): 299-310.

FOR ON-LINE PUBLICATION

Appendix to Kim, Putterman, and Zhang, 2019

“Trust, Beliefs, and Cooperation: An Experiment”

Overview

Appendix 1: A model of a psychological game

Appendix 2: Instructions

Appendix 3: Screenshots of the five environments and end-of-session survey

Appendix 4: Additional Tables and Figures

Appendix 1: A model of a psychological game

We provide a psychological game framework similar to that in Dufwenberg et al. (2011) for investigating the role of beliefs in contributions in the VCM.³⁸ In Dufwenberg *et al.* as well as Charness and Dufwenberg (2006), it is argued that others’ beliefs about what one will oneself choose to do may affect one’s behavior because of guilt or disappointment aversion: one does not want to harm another by disappointing what they expect of you, especially an expectation (belief) that one played a role in fostering. We refer to a subject’s belief about what others will do as her first-order belief, and we call her belief about what others believe her second-order belief. Hypothesis 5, which was derived from our framework, is consistent with Dufwenberg *et al.*’s discussion about the possible impact of first-order beliefs. As for second order beliefs, subjects in our experiment have reason to believe that others’ beliefs are influenced in a manner similar to their own by the information all alike receive about their five VCM environments.

A1.1. First order belief and contribution

Since strategic uncertainty is embedded in a situation where subjects make contributing decisions simultaneously in the VCM, one way of modelling conditional cooperation (the inclination to cooperate if others also cooperate) is to relate one’s first order beliefs to her contributing behavior. Following the formulation of Dufwenberg et al. (2011), the utility of player i is given as:

³⁸ Characterizing all equilibrium outcomes is not our primary purpose here because it is highly likely that there exist multiple equilibrium and then, it becomes a matter of equilibrium selection. Rather, we propose possible relationships between beliefs and contributions and given the degree of freedom in a model, it is easy to construct an equilibrium to support such outcomes.

$$u_i(a_i, a_j, b_{ij}) = 20 - a_i + (0.4) \sum_{i \in N} a_i + Y_i \sum_{j \in N \setminus \{i\}} \kappa_{ij} \lambda_{iji},$$

where N is a set of players belonging to the same environment and Y_i is a coefficient measuring player i 's sensitivity to kindness motivation. Suppose that κ_{ij} is player i 's kindness to j and λ_{iji} is player i 's beliefs about player j 's kindness to i . In the VCM context, j 's kindness towards i would be manifested as j 's contribution to the group account, so i 's belief about j 's kindness is a function of i 's belief about what j will contribute. Although it is a bit arbitrary, we will assume a reference for kindness as e which applies to all environments in the VCM.³⁹ Let's denote by b_{ij} player i 's beliefs about player j 's contribution, a_j .⁴⁰ By replacing κ_{ij} and λ_{iji} with $a_i - 10$ and $b_{ij} - 10$, we can rewrite the equation above as follows:

$$\begin{aligned} u_i(a_i, a_j, b_{ij}) &= 20 - a_i + (0.4) \sum_{i \in N} a_i + Y_i \sum_{j \in N \setminus \{i\}} (a_i - e)(b_{ij} - e) \\ &= 20 - a_i + (0.4) \sum_{i \in N} a_i + Y_i(a_i - e) \left[\sum_{j \in N \setminus \{i\}} b_{ij} - e \right]. \end{aligned}$$

In our experiment, being informed about others' trust or trustworthiness affects one's first order belief b_{ij} regarding what others will contribute (indexed by j), rendering b_{ij} higher (lower) in the high (low) trust of high (low) trustworthiness environment. Such differing first order beliefs will lead to different levels of contributing when one maximizes one's own utility.⁴¹ Therefore, this formulation can predict that there will be positive correlations between subjects' first order beliefs and contributions within and across environments.

A2.2. Second order beliefs and contribution

Guilt aversion is based on the idea that people don't want to let down others' expectations of their own behavior. Let's denote by c_{iji} player i 's belief about player j 's belief about player i 's contribution, i.e., player i 's belief about b_{ji} . Guilt aversion in the VCM can be captured by the notion that if a player contributes less than what the other group members believed he would contribute, he will experience disutility from letting down others' beliefs. Then, the utility function of player i can be presented as:

³⁹ For instance, one can have a reference point as the average of full contribution ($=20$) and no contribution ($=0$).

⁴⁰ In the interest of simplicity and symmetry with notations b and c as used in Dufwenberg et al. and this appendix, we here use a_j for the contribution of an individual, e.g. j , which is represented by C_j etc. in the text.

⁴¹ One extreme case is that if player i thinks, on average, the other 4 groups members are not kind ($\sum_{j \in N \setminus \{i\}} b_{ji} - e < 0$), her utility can be maximized by having $a_i = 0$. Otherwise, her utility can be maximized by $a_i = 20$ if Y_i is high enough and by $a_i = 0$ if Y_i is low enough.

$$u_i(a_i, a_j, c_{iji}) = 20 - a_i + (0.4) \sum_{i \in N} a_i - \gamma_i \max \left\{ 0, \frac{\sum_{j \in N \setminus \{i\}} c_{iji}}{4} - a_i \right\},$$

where $\gamma_i \geq 0$ is a coefficient measuring how sensitive player i is to guilt aversion. In an equilibrium, if γ_i is high enough, then conforming to others' second order beliefs can be a best response. As in the case of first order beliefs, the information regarding the level of trust or trustworthiness across environments will impact subjects' second order beliefs, and one can predict that there will be positive correlations between subjects' second order beliefs (which are positively impacted by the reported relative trusting or trustworthiness of those in the environment) and contributions across environments.

Appendix 2: Instructions

I. Instruction for Phase I

I.1 General Description

The interaction in this phase involves two roles that we refer to as the “first mover” and the “second mover.” The participants in each role are both endowed with 50 tokens at the beginning of the interaction. The first mover chooses how many token to send to the second mover, in 5 token increments, i.e. the first mover can send 0, 5, ..., or 50 tokens. **Denote this first mover decision as X .** Any tokens that are sent to the second mover by the first will be **tripled**. Upon receiving the tripled number of tokens, the second mover chooses how many tokens in integer amounts, from 0 to $3X$, to return to the first mover. **Denote this second mover decision as Y .**

Based on these decisions, the earnings of the first mover will be the initial 50 tokens minus the tokens (if any) he or she sent to the second mover plus the tokens (if any) that are returned by the second mover. The earnings of the second mover will be the initial 50 tokens plus three times the tokens (if any) that are sent from the first mover minus the tokens (if any) that are sent to the first mover. Namely, the payoff functions for the first mover and the second mover are, $\pi_1 = 50 - X + Y$ and $\pi_2 = 50 + 3X - Y$, respectively.

Example:

If the first mover, A, sent 25 tokens out of his endowment of 50, then the second mover, B, would receive $25 \times 3 = 75$ tokens in addition to her endowment of 50. Now suppose B decides to return 15 tokens, then A will end up with $50 - 25 + 15 = 40$ tokens, and B with $50 + 75 - 15 = 110$ tokens.

Remember that the second mover is free to return 0; likewise, the first mover is free to send 0.

Each of you will be making decisions first as a first mover, and then as a second mover. Only one of these roles can be selected for payment, and for whichever role that is, the computer will randomly select a participant to pair you with, so you will never play the second mover part against yourself as first mover. You will not learn the identity of the individual you are paired with, nor will they learn yours. Calculation of earnings will be explained later.

I.2 Decision as the “First Mover”

For this decision, you are the first mover. You can choose one of the 11 possible levels to send: 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50. Remember that any tokens you keep will be part of your own earnings if your choice in this role is selected to be paid off on, and every token you choose to send to the second mover is multiplied by 3.

You are allowed to experiment first by choosing one out of the 11 circles (corresponding to the 11 levels) on the screen (see below) and clicking on the “Calculate” button to see the selected numbers of tokens and see how many tokens the second mover would receive under this decision. Once you are satisfied with your choice, click “Submit” to confirm your decision.

Your decision is then final.

Phase 1: Decision 1

1. In this phase, you will be asked to make two sets of decisions. One of these two sets of decisions will be selected randomly for your actual payments at the end of the experiment.
2. You will now make the first of the two decisions. For this decision, you will be randomly paired with another participant. You are the first mover and the participant who you will be matched with is referred to as the second mover.
3. Enter your choice by clicking one of the circles, which are in 5 token increments from 0 to 50.
4. By clicking "Calculate," you can confirm the number of tokens associated with the selected circle and see how many tokens the second mover would receive. When you are satisfied with your decision, click "Submit". Otherwise, click "Back" to revise your decision.

- How much do you wish to send? : 0 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ 50

- Amount that will be sent to the second mover : 0

- Amount that will be received by the second mover : 0

I.3 Decision as the “Second Mover”

For this decision, you are the second mover. Please decide on how many (if any) tokens you choose to return to the first mover under all possible contingencies (see the screen below), as you will NOT be informed of the first mover's actual choice until payoffs are reported to you later.

That is, for each of the 10 relevant sent and tripled amounts (remember that the first mover can only send multiples of 5, up to 50, and if he sent 0, there is nothing you could return), please indicate how many tokens you choose to send back to the first mover, where you can send any integer amount between 0 and the amount you received. Please type your decision conditional on each received level in each of the ten boxes. You can play around with your answers and click “Calculate” to see your earnings and the first mover’s earnings resulting from your actions in each case. Once you are satisfied with your choice, click “Submit” to confirm your decision.

Your decisions are then final.

Phase 1: Decision 2

4. By clicking "Calculate", for each case, you confirm the number of tokens that you wish to send to the first mover and see how many tokens are allocated. When you are satisfied with your decisions, click "Submit". Otherwise, click "Back" to revise your decisions.

Sent (tokens)	Received (tokens)	Your decision (tokens returned to the first mover)	Your earnings (tokens)	First mover's earnings (tokens)
0	0	0	50	50
5	15	<input style="width: 80px;" type="text"/>	0	0
10	30	<input style="width: 80px;" type="text"/>	0	0
15	45	<input style="width: 80px;" type="text"/>	0	0
20	60	<input style="width: 80px;" type="text"/>	0	0
25	75	<input style="width: 80px;" type="text"/>	0	0
30	90	<input style="width: 80px;" type="text"/>	0	0
35	105	<input style="width: 80px;" type="text"/>	0	0
40	120	<input style="width: 80px;" type="text"/>	0	0
45	135	<input style="width: 80px;" type="text"/>	0	0
50	150	<input style="width: 80px;" type="text"/>	0	0

I.4 Payoff Calculation in this Phase

Only one of the two decisions you made will be randomly selected to determine your earnings in this phase. If your first mover decision is selected for payment, your decision and the second mover decision of the participant randomly paired with you will determine your and his/her earnings in this phase. Likewise, if your second mover decision is selected for payment, your decision and that of the first mover randomly paired with you will determine your and his/her earnings in this phase. Your earnings this phase in tokens will convert to real money at the rate of \$0.10 per token, which will be paid to you at the end of the experiment. (Note that your counterpart's decision and your resulting earnings in this phase will not be reported to you until after Phase 2.)

Example:

If Mr. A's first mover decision is chosen for his payment, and he got 60 tokens as the first mover in the anonymous pairing, then he will receive $60 * \$0.10 = \6 at the end of the experiment as his earnings for this phase.

*This is the end of the Instructions for Phase I.
Please raise your hand if you have any questions.*

II. Instruction for Phase II

II.0 Brief Introduction

In this phase, you are asked to make 3 sets of decisions. Each set needs your decisions in five different decision environments, where you have different information about the other participants matched with you. The first set of decisions are about allocating tokens, while the second and third sets are about estimates of what others will do and guesses about others' estimates. Group formation within each environment is random (although in some environments subject to relevant constraints), and no information about the other participants' identities or decisions regarding allocation and estimation will be revealed to you or others throughout the phase. At the end of the experiment, you will only be notified of others' allocation decisions for the purpose of payment, and only **one** environment will be chosen for payment. Payoff calculations in this phase will be explained later.

II.1 The Five Decision Environments

Within each set of decision, there are five different environments that differ in terms of the (true) information about the group composition.

In **Environment 1**: the other four group members are chosen randomly from among all participants.

The differences across the remaining four environments are based on differences in participants' choices in Phase 1 (the first mover/second mover decision).

To understand environments 2 and 3, suppose that all phase 1 first-mover sending decisions in this room are ranked from the lowest (1st) to the highest (20th) to form an ordered list. Based on such a list,

In **Environment 2**: the average level of the other four members' sending corresponds to a low rank (below rank number 8).

In **Environment 3**: the average level of the other four members' sending corresponds to a high rank (above rank number 12).

Put more intuitively (but a bit less precisely), in Environment 2 you are grouped with others who on average sent relatively small amounts as 1st movers, while in Environment 3 you are grouped with others who on average sent relatively large amounts as 1st movers.

To understand environments 4 and 5, suppose that all Phase 1 second-mover returning decisions, when being sent all 50 tokens by the first mover, are ranked from the lowest (1st) to the highest (20th) to form an ordered list. Based on such a list,

In **Environment 4**: the average level of the other four members' returning corresponds to a low rank (below rank number 8).

In **Environment 5**: the average level of the other four members' returning corresponds to a high rank (above rank number 12).

Put more intuitively (but a bit less precisely), in Environment 4 you are grouped with others who on average returned relatively small proportions as 2nd movers, while in Environment 5 you are grouped with others who on average returned relatively large proportions as 2nd movers.

II.2 The Three sets of Decisions

Decision 1:

In this decision set, you will be a member of a group consisting of **5 people**, yourself included, in all five different environments. As explained in the previous section, the other members of your group differ in each environment. Note that this first decision is the most payoff-relevant stage in this phase; that is, it will probably determine the largest part of your payment for the phase.

In all environments, each of you is endowed with 20 tokens at the beginning and each simultaneously makes individual decisions on how to allocate these tokens, in integer amount, between a group account and a private account. Any tokens you choose not to allocate to the group account will be automatically allocated to your private account. Everyone benefits equally from the tokens in the group account: each of you gets 0.4 tokens towards your private account per token in the group account. That is, your earnings are the number of tokens in your private account plus 0.4*the total tokens in the group account.

Example:

	CASE 1	CASE 2	CASE 3	CASE 4
a. Your Contribution	0	20	10	5
b. Other's contribution - 1	0	20	10	6
c. Other's contribution - 2	0	20	10	7
d. Other's contribution - 3	0	20	10	13
e. Other's contribution - 4	0	20	10	9
f. Tokens in your private account (= 20 – a.)	20	0	10	15
g. Tokens in the group account (= a+b+c+d+e)	0	100	50	40
h. Earnings from group account for each person (=0.4*g)	0	40	20	16
Your Total Earnings (in tokens) (= f + h)	20	40	30	31

Decision 2:

In this decision set, you will be asked to estimate, **on average**, how many tokens the other four group members have allocated to the group account in Decision 1 in each of the five environments. 5 additional tokens will be given to you if your estimate is within one token of the true average in the payoff relevant environment. No tokens will be taken away from you if your estimate is incorrect/ imprecise.

Example:

1) Suppose in Environment 1, the actual average allocation of the other four people in your Environment 1 group is 5. If this environment is selected for payment, then if your estimate of

the average is between 4 and 6 tokens, you will get 5 additional tokens; otherwise you will NOT get any additional tokens

2) Suppose your guess for Environment 1 qualifies you for the 5 additional tokens, while your estimate for Environment 2 does not. If Environment 2 is chosen for payment, then you will NOT get any additional tokens for this decision.

Decision 3:

In this decision set, you will be asked to estimate the **average estimate** provided by the other four group members in Decision 2. That is, you will guess the average of the estimates that each of the other four members has given regarding the average allocation to the group account, by you and the rest, in the previous decision. Similar to Decision 2, you will receive an additional 5 tokens if your estimate comes within one token of the true average of their estimates in the payoff relevant environment. No tokens will be taken away from you if your estimate is incorrect/ imprecise.

II.3 Payoff Calculation for this Phase & the entire Experiment

For Phase 2, only one out of the five decision environments will be randomly chosen to determine your earnings for both your allocation decision (Decision 1) and your estimates (Decisions 2 and 3). For the selected environment, your earnings in tokens will be based on Decision 1 (i.e., what you put in your private account plus the earnings from the public account based on what you and the other four in that group put in the public account), plus any additional earnings from estimates in Decision 2 and Decision 3. Your token earnings will be converted to real money at a rate of **\$0.20 per token**.

Example:

Suppose Environment 1 is chosen for payment, and your estimate in Decision 2 is in the correct range but that in Decision 3 is not; moreover, you contributed 10 tokens in Decision 1 and the public account ends up with 45 tokens. In the end, you would get: $\{(20-10) + 0.4*45\} + 5 + 0 = 33$ tokens, which is $33*\$0.20 = \6.60 in real dollars, for this phase.

In addition to knowing the results in phase 2, you will also be given information on results in Phase 1 (refer back to page 4 of the first set of instructions). Your final earnings from the experimental decisions and outcomes are then the sum of your earnings in Phase 1 and Phase 2. Your full payment will be this sum plus the \$5 show up fee.

***This is the end of the instructions for Phase 2.
Please raise your hand if you have any questions.***

Appendix 3: Screenshots of the five environments and end-of-session survey

Phase 2: Decision 1

1. In this phase, you will be asked to make three sets of decisions, making each decision once for each of 5 environments.

2. You will be a member of a group consisting of 5 people including yourself. Each group member has to decide on the allocation of 20 tokens.
You can put none, some, or all of these 20 tokens into a public account. Each token you do not put into the public account will automatically remain in your private account.

Please enter the amount you wish to allocate to the public account for Environment 1
(group members are assigned randomly)

Submit

Phase 2: Decision 1 continued

Please enter the amount you wish to allocate to the public account for the environments 2 and 3.

Enter your allocation to the public account in Environment 2
(other group members' average rank as 1st mover senders is low
[< 8th])

Enter your allocation to the public account in Environment 3
(other group members' average rank as 1st mover senders is high
[> 12th])

Submit

Phase 2: Decision 1 continued

Please enter the amount you wish to allocate to the public account for the environments 4 and 5.

Enter your allocation to the public account in Environment 4
(other group members' average rank as 2nd mover returners is low
[< 8th])

Enter your allocation to the public account in Environment 5
(other group members' average rank as 2nd mover returners is high
[> 12th])

Submit

Phase 2: Decision 2

1. In the previous decision, you and the other four members in each of five possible groups decided how many tokens they allocated to the public account.

2. For this decision, you will be asked about your estimate of how many tokens the other four members in each of the five possible groups allocated to the public account, on average.

Please enter your estimate for Environment 1
(group members are assigned randomly)

Submit

Phase 2: Decision 2 continued

Please enter your estimate of the average number of tokens that the other four members of the group allocated to the public account for the environments 2 and 3.

Enter your estimate for Environment 2
(other group members' average rank as 1st mover senders is low
[< 8th])

Enter your estimate for Environment 3
(other group members' average rank as 1st mover senders is high
[> 12th])

Submit

Phase 2: Decision 2 continued

Please enter your estimate of the average number of tokens that the other four members of the group allocated to the public account for the environments 4 and 5.

Enter your estimate for Environment 4
(other group members' average rank as 2nd mover returners is low
[< 8th])

Enter your estimate for Environment 5
(other group members' average rank as 2nd mover returners is high
[> 12th])

Submit

Phase 2: Decision 3

1. In the previous decision, you and the other four members in each of five possible groups estimated the average number of tokens that the other four members allocated to the public account.

2. For this decision, you are asked to give your best estimate of what average estimate the other four members in the group themselves gave; that is, you will guess or estimate what their estimates were, on average.

Please enter your estimate for Environment 1
(group members are assigned randomly)

Submit

Phase 2: Decision 3 continued

Please enter your estimate of the average estimate the other four members of the group
stated in the previous decision for the environments 2 and 3.

Enter your estimate for Environment 2
(other group members' average rank as 1st mover senders is low
[< 8th])

Enter your estimate for Environment 3
(other group members' average rank as 1st mover senders is high
[> 12th])

Submit

Phase 2: Decision 3 continued

Please enter your estimate of the average estimate the other four members of the group stated in the previous decision for the environments 4 and 5.

Enter your estimate for Environment 4
(other group members' average rank as 2nd mover returners is low
[< 8th])

Enter your estimate for Environment 5
(other group members' average rank as 2nd mover returners is high
[> 12th])

Submit

What is your class status?

- Answer:
- ☐ Freshman
 - ☐ Sophomore
 - ☐ Junior
 - ☐ Senior

continue

Which ethnicity origin (or race) describes you the best?

Answer:

- ☐ White
- ☐ Hispanic or Latino
- ☐ Black or African American
- ☐ Native American or American Indian
- ☐ Asian / Pacific Islander
- ☐ Other

continue

We are interested in understanding the thinking behind your decisions in today's experiment.
In phase 1, on what basis did you decide how much to send as the first mover or to return as the second mover?

Answer:

continue

We are interested in understanding the thinking behind your decisions in today's experiment.
In phase 2, on what basis did you decide how much to allocate to the public account in different environments?

Answer:

continue

During the experiment, you considered the likely decisions of those who sent relatively small and relatively large amounts as first movers, and of those who returned relatively small and relatively large amounts as second movers. We'd like to know whether you think that high first move senders are likely to also be high second move returners, and vice versa. How highly do you think these two decision tendencies are correlated, for given individuals?

Answer:

- ☐ Almost perfectly correlated
- ☐ Highly correlated
- ☐ Mildly correlated
- ☐ Not correlated
- ☐ Negatively correlated (high first mover sending is associated with low second mover returning)

continue

To what extent did the other group members' decisions conform with your estimates, and to what extent and in what ways did their decisions surprise you?

Answer:

continue

If you have any suggestion for this experiment, please let us know!! Any comment or suggestion will be appreciated!

Answer:

continue

Online Appendix 4: Supplementary Regression Table

Table-A1 Contribution and Beliefs in the VCM (2 step verification)

Dependent Variable:	(1)	(2)	(3)	(4)	(5)
Contributions	Random	Low Trust	High Trust	Low Trustworthiness	High Trustworthiness
First-Order Beliefs	1.032*** (0.0786)	1.016*** (0.146)	0.936*** (0.0850)	1.222*** (0.0853)	1.044*** (0.0681)
Residuals	-2.81e-09 (0.268)	5.16e-09 (0.0956)	-7.75e-10 (0.206)	4.44e-09 (0.159)	1.62e-09 (0.234)
Constant	2.745 (1.964)	1.483 (2.732)	0.346 (2.431)	2.016 (2.325)	-0.521 (2.106)
Demographic Controls	YES	YES	YES	YES	YES
Observations	120	120	120	120	120
R-squared	0.580	0.544	0.476	0.638	0.543

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1