Trust, Beliefs and Cooperation:
Excavating a Foundation of Strong Economies

Jeongbin Kim, Louis Putterman and Xinyi Zhang*

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 indicate 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.

Keywords: trust, cooperation, beliefs, reciprocity, economic growth
JEL codes: D01, C91, C92, H41

* 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, Jack Fanning, Xu Zhang, Hui Wen Ng, Oleg Semenov and other participants in the theory workshop at Brown University for helpful comments and suggestions, and we are grateful to Jackson Cline for able research assistance.

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 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 establishing and maintaining 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.

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 (because we announce truthfully that ties are broken randomly), 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. 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 highly 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 returned (or sent) 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 paper 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. Social welfare rises when first-movers make such transfers in justified anticipation of such trustworthiness, an action or state of mind that many researchers subsequently denoted trust (e.g., Fehr, 2009; Eckel and Wilson, 2004). Assuming that preferences such as altruism are not important motivations behind the act of the first movers (Cox,

2004; Ashraf *et al*., 2006), 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) and the possibility of achieving higher payoffs (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 belief in the trustworthiness of others is correlated with one's own trustworthiness since it is obtained partly by introspection (Sapienza *et al*., 2013).

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 the kinds of reputational considerations that the standard experimental design rules out. On the other hand, as considered next, second mover returning 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] The Berg *et al*. design made it likely that first-mover sending was not altruistically motivated because first and second movers were recruited from a common subject pool and were each provided with initial endowments of $10. There may, of course, be 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.

[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.

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; Sobel, 2005; 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 disposition, including Gächter and Thöni (2005), Gunnthorsdottir *et al*. (2007), and Ones and

[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 a stag hunt game from the standpoint of the subjectively evaluated payoffs.

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 subjects' survey-measured or lab-measured trust, or both, are associated with their contribution behaviors 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 not in identifying how individuals' own behaviors extend across domains, but rather in showing how beliefs about the trust and trustworthiness *of others*, beliefs based partly on information conveyed by us but also crucially on the beliefs subjects bring to the lab, shape beliefs about others' cooperativeness and how those beliefs then 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.

[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.

2.3. Manipulating beliefs

Our paper can be situated among 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 whether a VCM is labeled a Community game or a Market game causes subjects to 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*. (2012) 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 own and others' beliefs, namely providing both oneself and others with information about 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)—in particular, 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 or of what the latter consists of, 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, but all feedback is withheld until after both phases of decision-making, shortly before distribution of earnings.

Everything described in the following sub-sections was common knowledge among all subjects. The instructions are included in the online Appendix. The experiment consists of a single treatment incorporating multiple decision conditions, 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.

Formally, payoffs for subject *i* as 1st mover (sender) can be written

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

and for subject *j* as 2nd mover:

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

where $a \in [0, 5, \ldots, 50]$, and $b \in [0, 1, 2, \ldots, 3a]$. Subjects were informed that each token would convert to \$0.10 of real money, payable at the end of the experiment.

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 receiver'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, as in much of the literature (Eckel and Wilson, 2004; Johnson and Mislin, 2011), we treat the amount sent as a measure of trusting. 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 membership in the other four environments was determined with the aid of rankings of first phase behaviors. Roughly speaking, groups were formed so that their members would be (respectively) 'relatively trusting,' 'relatively untrusting,' 'relatively trustworthy,' and 'relatively untrustworthy' participants, on average. 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 we assigned subjects numerical ranks 1 – 20 for lowest to highest first mover sending, and

[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.

likewise for lowest to highest second mover return proportion, with ties broken randomly. This way, each subject was identified by two numbers, a 'trust rank' for her sending decision and a 'trustworthiness rank' for her returning decision. Because ties were broken randomly, these ranks would be informative to the degree that subjects differed in their Phase 1 decisions. Subjects were truthfully told that in environment 2 (in effect a 'high trust' group), 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 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 the 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:

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

[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. Moreover, we later demonstrate that any effect of this factor on beliefs was more than offset by a countervailing phenomenon (see end of section 5.2, and fn. 29).

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 anything. Each token earned in this phase of the experiment 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. 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 completion of P. 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, and his/her earnings. 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_i. For some individuals, this disposition is very low and does not impact their behavior. For others, it is stronger and plays an important role both in driving trustworthiness and trusting in the trust game, and in driving conditional cooperation in the VCM. In a strategic interaction, considerable attention should be given to how people form beliefs about others' reciprocity.

We first present a simple model with a particular emphasis on reciprocity, and then discuss its implications for our experimental design. Assume a trust game second-mover *j* having underlying reciprocity disposition r_j. Ignoring additional factors that might influence

[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.

[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.

trustworthiness, we focus on how the disposition r_j affects *j*'s trustworthiness (the proportion of receipts which she returns) in that game, denoted by $r_t(r_j)$. Assume that a second mover *j*'s utility function is defined as follows:

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

where x_i is the amount that first mover *i* sent, y_i is 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 her reciprocity.[13] One might view $r_t(r_j) \cdot 3x_i$ as *j*'s benchmark conception of how much she must send back according to her ideal of reciprocity. If $\mu_j > 1$, her utility gain from closing the gap between y_j and that ideal suffices to offset her utility loss from forfeiting the associated monetary payoff and y_j becomes a function of r_j and x_i.[14] Note that this model is identical to the standard approach if $\mu_j < 1$ or if $r_t(r_j) = 0$ for all *j*.[15]

Next, we model first mover sending. Since first mover *i*'s own reciprocity has no bearing on her decision and our model abstracts from other considerations, we assume that *i* simply maximizes

$$U_i = 50 - x_i + y_j.^{16} \qquad (5)$$

In a one-shot interaction like our experiment, there exists strategic uncertainty, and a first mover *i* needs to form a belief or estimate, y_j^{ei}, about second mover *j*'s return amount, y_j. Given that y_j is a function of *j*'s reciprocity r_j, we can write $y_j^{ei} = y_j(x_i, r_j^{ei})$, indicating that *i*'s belief about *j*'s

[13] The utility function above can be regarded as an application of Kőszegi and Rabin (2006)'s reference dependence model of 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, the strategy 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 disposition which will be instrumentally used in the analysis that follows.

[14] In a more detailed analysis, the fraction of the tripled amount returned might be increasing in the amount of trust implied by *i*'s chosen x_i. Function r_t might also 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, 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.

[15] This implies that given a positive amount of sending, a zero return decision by a second-mover could be due either to $\mu_j = 0$ or to $r_t(r_j) = 0$. If we observe a positive amount of returning, in contrast, we can infer $\mu_j > 1$.

[16] Note that each subject in our experiment makes both first-mover and second-mover decisions, so our modeling of first-mover choice as if made by a simple money payoff maximizer is in no way equivalent to assuming that there exists a subset of agents (say, a type *i*) that are devoid of social preferences. We merely simplify by assigning r no direct role in the first-mover decisions. We in fact show it to play a strong indirect role in those decisions, in our subsequent analysis.

reciprocity plays a key role in determining the amount sent. 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 $r_t(r_j^{ei}) < ⅓$, and $x_i = 50$ for r_j such that $r_t(r_j^{ei}) \geq ⅓$.[17] 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_t that assign positive probabilities to a range of values, in consequence of which first-mover sending can also take intermediate values (explicit modification of (5) is omitted for the sake of brevity). We accordingly expect x_i to be an increasing function of the first mover's belief about r_j, without necessarily having an abrupt switch point.

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

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

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 $r_c'(r_i) \geq 0$ is the rate at which *i* matches the expected contributions of other group members.[18] In our experiment, each subject is asked to report $\sum c_{-i}^{ei}/(\text{n} - 1)$, her 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. That is, 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, with neither interaction partner needing to form a belief about the other's belief. In the VCM, beliefs about beliefs may influence agents' decisions as each of five symmetrically situated members tries to select a contribution c_i.[19] In one extreme case, if $r_i = 0$ (or $\gamma_i < 1$), subject *i* will contribute nothing to the group account as in the case with standard preferences. However, a subject having $\gamma_i > 1$ and $r_i > 0$ would contribute $c_i = \frac{r_c(r_i)}{n-1}\sum c_{-i}^{ei} > 0$. That is, a subject with high r_i is likely to condition his/her contribution on beliefs about others' average contributions, i.e., to be a conditional cooperator. In this case, own

[17] Alternatively, x_i could take any value when *j* is believed to return exactly ⅓, but we think it implausible that this knife-edge possibility is what accounts for most observed intermediate choices. Note also that rather than characterizing all equilibria, our discussion focuses on subjects' best responses to their beliefs about others' reciprocity, and on the consequence of own reciprocity on behavior.

[18] We assume that γ_j is highly correlated with μ_j, meaning that a subject who places a high weight on reciprocity in the trust game also gives considerable weight to reciprocity in the VCM.

[19] 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 in question into an estimate of the distribution of $[\sum c_{-i}/(\text{n} - 1)]$. We reiterate, as in footnote 17, that we focus on best responses to beliefs rather than on equilibria of beliefs. The substantial power of introspection bias manifest in our data is consistent with the possibility that much action occurs—even in the real world—outside of full expectations equilibrium, perhaps because agents encounter many different game forms no one of which is played sufficiently often.

contribution is a function of both own reciprocity and beliefs about others' average contributions, which are in turn related to beliefs about their beliefs and reciprocity (c_i is a function of r_i, and hence, c_{-i}^{ei} is a function of r_{-i}^{ei}). In consequence, we can expect to see positive relationships between first-order beliefs and contributions for subjects with high r.[20]

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. How subjects form beliefs of what other group members will contribute to the group account is essential, and beliefs about others' contributions are non-decreasing in beliefs about their r's. Relatedly, 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 lay out specific hypotheses based on this discussion in the following sub-section.

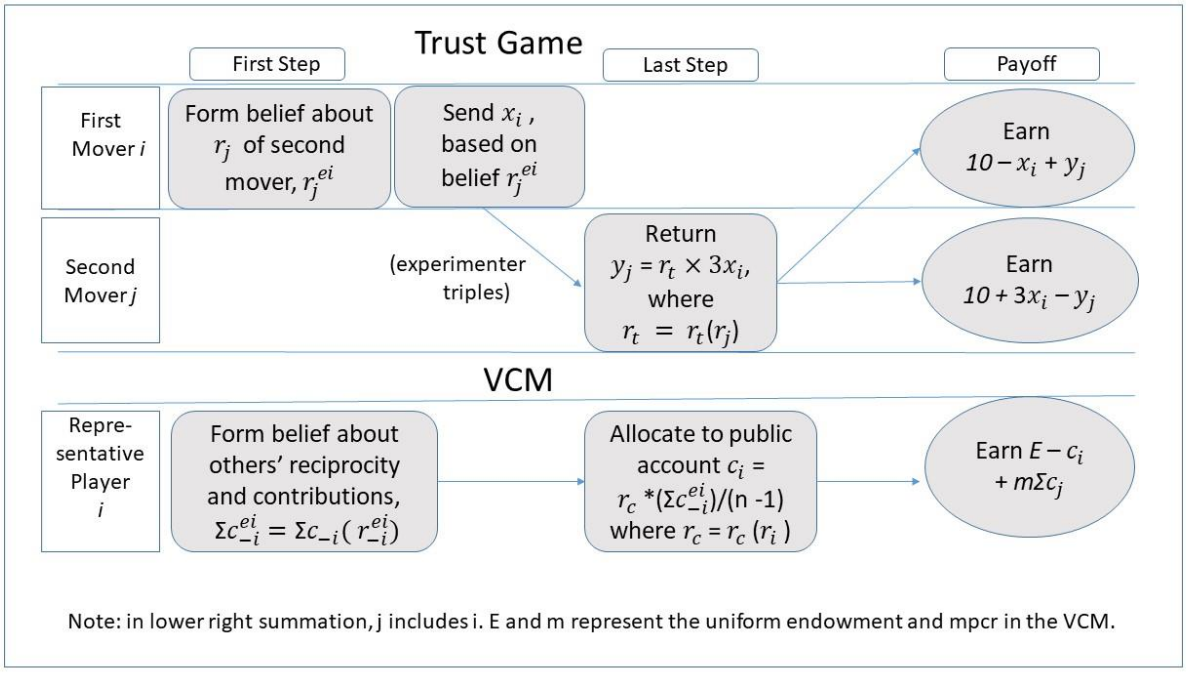


Figure 1. Diagrammatic representation of trust game and VCM.

[20] Qualitatively similar predictions can be drawn by using a psychological game framework (Geanakoplos *et al*, 1989). For instance, Dufwenberg *et al.* (2011) use a psychological game to model associations of first- and second-order beliefs with contributions in the VCM, and like us, they focus on individuals' responses to their beliefs rather than on equilibrium beliefs (see e.g. p. 463).

We first further clarify the similarities and differences between the trust game and VCM with the help of Figure 1, which summarizes schematically how the VCM collapses the two trust game roles into one, while identifying 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 (assuming $x_i > 0$) and VCM contribution c_i (assuming $\sum c_{-i}^{ei} > 0$) both being higher the higher is own r, *ceteris paribus* (and assuming γ and $\mu > 1$). 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 assumed by 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 externally 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 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 slightly 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 considerably affected by own r. This is so despite 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 formed group environment, being likely to be shared by a great many subjects. 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 one step further removed from own r_i, according to our framework. One of the most interesting features of our data will be what they tell us about subjects' assessments of how indicative trustworthiness is of trusting.

4.2. Hypotheses

We next lay out our hypotheses regarding behaviors in our experiment using the framework above plus three auxiliary assumptions. The latter are, first, an assumption that the level of reciprocity varies within any given population. Second, we assume that most participants believe that the level of reciprocity varies within our population of subjects. Third, we assume that as a consequence of introspection bias, beliefs regarding the average level of reciprocity in the population vary among subjects in a manner correlated with their own levels 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_i is higher than is *k*'s when provided with the same incomplete information about that population. We can now turn to the hypotheses themselves, the first of which tests our assumption of introspection bias.

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.*

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

Hypothesis 2 (aggregate first-order beliefs and group formation rules): **(a)** *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.* **(b)** *Subjects who as trust game second movers return relatively large proportions have higher first-order beliefs in each given environment in the VCM than do subjects with lower return proportions.*

H2(a) follows from the framework laid out in conjunction with Figure 1 and our assumption in this sub-section that most subjects believe that r_i varies within the population. The belief that r_i 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_i, with average r_i being of intermediate value in a randomly formed group. Since higher (lower) r_i also predicts higher VCM contribution, according to our framework, the estimate of $[\Sigma c_{-i}/(n-1)]$ will vary as indicated. H2(b) follows because assessment of others' likelihoods of reciprocity is also affected by introspection bias.[21]

[21] The tendency predicted in H2(b) should be countered, for highly rational subjects, by the factor discussed in footnote 10 regarding selection of other group members needing to differ depending on own behavior in order to meet the stated selection criteria. Our adoption of H2(b) as our prediction reflects our expectation that such rational calculations are likely to be rare and overshadowed by the force of introspection bias.

Our most important prediction concerns the relationship between contributions and first-order beliefs in the VCM. We predict that at the aggregate level, the differences in contributions across high and low environments will be related to the differences in first-order beliefs in the corresponding environments because on average our subject population will conditionally cooperate—i.e., subjects on average contribute more the more others in their group are believed to contribute. We note, however, that the effect of belief on own contribution is by no means uniform. Although subjects having high r_c are expected to positively adjust their VCM contribution according to their belief about other group members, which varies by environment, subjects with low r_c for whom $r_c = 0$ are expected not to contribute regardless of environment. H3 spells out the implication of H2 assuming the average subject to be conditionally cooperative:

Hypothesis 3 (first-order beliefs and contributions) *Subjects on average contribute more (less) to the group account when in high (low) environments than when in a random environment.*

H4 spells out how variation in trustworthiness predicts variation in conditional cooperation type:

Hypothesis 4 (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 less (or nothing), who in turn are more likely to be free riders in the VCM.*

Together, H2(b), H3 and H4 imply that more trustworthy subjects will have higher first-order beliefs and will contribute more in the VCM. But given the posited link between trust and trustworthiness due to introspection (H1), the same statements regarding correlations with first order beliefs and VCM contributions can also be applied to trust. Although the relationship of trustworthiness to each of the two variables is more direct than is that of trust, we anticipate the possibility of correlations for trust also, and we will therefor run econometric "horse races" to determine empirically whether trust or trustworthiness have stronger predictive power for contributions in the VCM, or whether the associations for both are equally strong. Regardless of which variable proves to have higher predictive power, we anticipate that each will show significant correlations in its own right (at a minimum, when the other is not controlled for). This implies:

Hypothesis 5 (trust and trustworthiness in the trust game and contributions in the VCM): *Contributions in the VCM will be higher (lower) for subjects exhibiting higher (lower) trust and trustworthiness in the trust game.*

Although our discussion has emphasized the relationship between first-order beliefs and contributions 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 also be influenced by their beliefs regarding others' beliefs, because they would incur guilt by letting those others' expectations down.

Hypothesis 6 (second-order beliefs, guilt aversion and VCM contributions): *Subjects have higher (lower) second-order beliefs about others' average VCM contribution beliefs 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.*

It is an empirical question which order of beliefs is a more prominent channel through which relative trustworthiness in the trust game influences contributing behavior in the VCM. We perform econometric testing to determine which order beliefs is more crucial for our subjects.

5. Results and Analysis

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, about 4.5% above the female share of the overall student body. Each session lasted around 1.5 hours, and average earnings were $18.20, consisting of a $5 show-up fee, Phase 1 earnings averaging $6.57, Phase 2 earnings averaging $6.52, and $0.11 for accuracy in belief elicitations.

5.1. Behavior in Trust Game

We begin by presenting a description of behavior in the trust game. On average, our subjects sent 19.6 tokens (39.2% of their endowments) as senders, 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 observed in previous experiments, albeit a little below average.[22] Our first 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

[22] 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 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").

mover sending (as percentage of the 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.[23]

Result 1. Individuals' second-mover behaviors in the trust game are highly correlated with their first-mover behaviors, consistent with H1 and the assumption of introspection bias.

5.2. First-order beliefs in the VCM

The next prediction of our framework concerns how first-order beliefs in the VCM are shaped by (1) the information about others' relative trust/trustworthiness and (2) own reciprocity type. H2(a) predicts a specific pattern of heterogeneity in first-order beliefs about other group members' average contributions across environments. Figure 2 depicts the average first-order belief in each environment. Numbers indicate the average first-order belief and error bars indicate the 95% confidence interval.

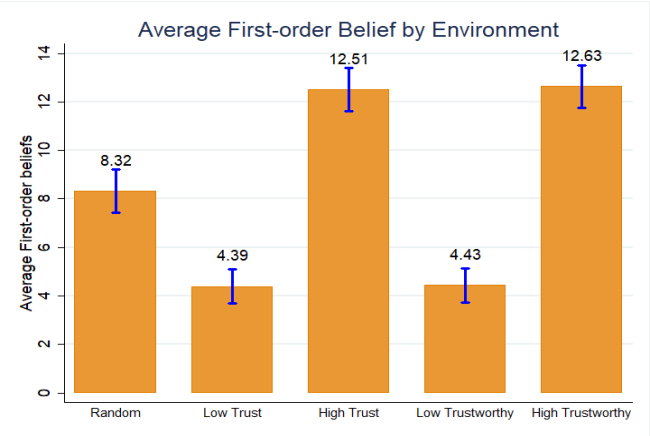


Figure 2. Average first-order beliefs about others' contributions in the VCM, by environment

Clearly, first-order beliefs about others' contributions in the VCM do vary by environment, and in the manner predicted, i.e. expected contributions of others are higher in the high environments and lower in the low environments than those in the random environment. All differences versus the random environment, and each difference between a high and a low environment, are highly significant (Mann-Whitney tests, p-values < 0.001), confirming H2(a).[24] This implies that the information about relative trusting/trustworthy behavior in the trust game strongly influences first-order belief formation in the VCM. Furthermore, while our conceptual framework views trustworthiness as more directly related to contributing than is trust, we find that the effects of the information conveyed by high (low) trust on the first-order beliefs is not distinguishable from that

[23] 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.

[24] We report two-sided tests, although the fact that we are testing explicit hypotheses would justify the use of one-sided tests, leading to even lower p-values (each half the size of those reported).

of high (low) trustworthiness (Mann-Whitney tests, p-values > 0.85). This finding would be consistent with a preponderance of subjects themselves believing there to be a high correlation of trust and trustworthiness via introspection bias, paralleling one of the assumptions we used in constructing our predictions in Section 4.

Result 2: (a) 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, being higher for high trust and high trustworthiness environments than for the random environment, in which it is in turn higher than in low trust and low trustworthiness environments, as predicted by H2(a). (b) The impact on beliefs of high (low) trustworthiness does not differ from the impact of high (low) trust.

H2(b) concerns whether, in addition to heterogeneity between environments, first-order beliefs differ also within given environments due to differences in the trustworthiness (and underlying that, in the r_j's) of the subjects forming those beliefs. It predicts that more (less) trustworthy subjects will have more (less) optimistic beliefs about the cooperativeness of others in a given environment as a consequence of introspection bias. It turns out to be straightforward to show that subjects' 2nd mover choices and beliefs in each VCM environment are highly significantly correlated. An OLS regression for each environment with robust standard errors clustered at the session level in which first-order VCM belief is regressed on individuals' average returning ratios and a constant yields a positive coefficient significant at the 1% level for each environment (see Table A.1). In parallel regressions substituting the proportion of 1st mover endowment that is sent in the trust game for the 2nd mover return share, the alternative explanatory variable also obtains positive coefficients, but ones significant in the low trust and low trustworthiness environment regressions only. Table 2 displays what might be considered a more demanding set of tests of H2(b): regressions for first-order beliefs in each environment that control simultaneously for 1st and 2nd mover behaviors in the trust game. The estimates show that in line with H2(b), trustworthiness (avg. returning %) is positively and at least marginally significantly correlated with first-order beliefs in all environments, with the partial correlation coefficients being significant at the 5% level in the low trust and high trustworthiness environments (columns (2) and (5)) and at the 1% level in the low trustworthiness environment (column (4)). The sending share also obtains positive coefficients, but they are significant at the 5% level in only one environment and marginally significant in one other. As an additional exercise, we pool the data from all environments but maintain the focus on within-environment differences by using as dependent variable the difference between subject *i*'s average first-order belief and the average first-order belief of all participants, in the given environment. The result, shown in column 6, again supports H2(b) with a highly significant coefficient on *i*'s return share, although *i*'s sending share is also quite significant in this estimate. Finally, when we calculate the average contribution of all groups of four others who were eligible to be in a given environment for each subject based on the subject's own rank for trusting and trustworthiness, we find that subjects ranking low for trustworthiness tended to systematically underestimate the contributions of those in each

environment, whereas those ranking high for trustworthiness tended to systematically overestimate the others' contributions. The ratio of belief about others' contribution in each selective environ-

Table 2. Decisions 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	2.364*	0.677	2.124**	1.236	1.495
in Trust Game	(1.923)	(0.948)	(1.698)	(0.657)	(2.219)	(1.257)
Average Returning %	2.447*	3.134**	4.576*	2.533***	4.873**	3.513**
in Trust Game	(1.062)	(0.779)	(1.988)	(0.614)	(1.691)	(1.009)
Constant	7.133***	2.488***	10.82***	2.811***	10.63***	-1.681**
	(0.679)	(0.446)	(0.742)	(0.340)	(0.895)	(0.546)
Observations	120	120	120	120	120	600
R-squared	0.029	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	0.0037

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

ment (environments 2 – 5) to eligible others' actual contributions in the relevant environment is significantly positively correlated with of the individual's own average return rank in the trust game, with correlations having *p* values between 0.001 and 0.02, depending on environment, and regression correlation coefficients significant at 1% or 5% levels (see Table A.2).[25] To summarize:

[25] Notice that two considerations could impact the relationship between behaviors of eligible group members, beliefs about those behaviors, and own trust game behaviors. As was noted in the design section, a highly trustworthy subject should anticipate somewhat lower trustworthiness of others in her high trustworthiness environment than should a very untrustworthy subject, since the other four must be more highly ranked for trustworthiness in order for the group of five to meet the high trustworthiness rank criterion on average, in the latter's case. Insofar as VCM contribution is predictable from trustworthiness, as our framework supposes, the less trustworthy subject should therefore anticipate higher contributions than should the more trustworthy one, in the respective high trustworthiness environments that they will encounter. Introspection bias points in the opposite direction, implying that more trustworthy subjects are more optimistic about others' reciprocal tendencies, ceteris paribus. There is no a priori way to predict which tendency would be stronger, but in H2(b) we predicted greater strength for introspection bias, due to our anticipation that their belief formation biases would be stronger than the powers of rational calculation for most subjects. The results strongly bear this out. We suspect that the rational calculus concern may in fact have crossed the minds of only a small minority of subjects, though it was noticed by several economic theorists in our seminar audiences.

Result 3: (a) Subjects who return more (less) in the trust game have significantly higher (lower) beliefs about how much others will contribute in our VCM environments, supporting H2(b). (b)The relationship between trustworthiness and first-order beliefs is stronger than that between trust and first-order beliefs.

5.3. Contributions in the VCM

Our third major prediction is about the relationship between contributions and first-order beliefs in the VCM. Given the effect of the information about trustworthiness on first-order beliefs across environments, we look at whether average contributions track average first-order belief levels closely in all environments. Figure 3 shows that in the VCM, 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.

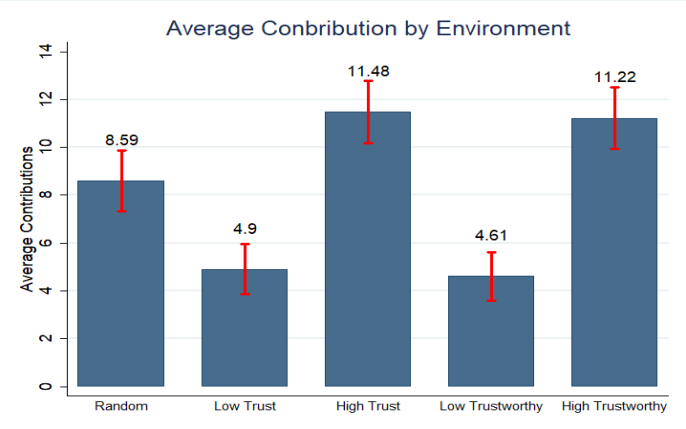


Figure 3. Average Contributions in the VCM, by environment

Compared to the contributions in the random environment, subjects contributed substantially more in both high trust and high trustworthiness environments and less in low trust and low trustworthiness environments, respectively, with all the differences significant (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 an ostensibly low trust/trustworthiness to an ostensibly high trust/trustworthiness environment. The differences in contribution between high, random and low environments indicate that the differing information about trust and trustworthiness rankings has large effects on contributing behavior in the VCM. The directionality of the observed differences is as would be predicted given the observed pattern of first-order beliefs about others' VCM contributions in each environment plus an assumption that on average subjects are conditionally cooperative (an assumption that will be shown to be in line with the evidence—see below).

In contrast to the significant differences in contributions between high and low (relative) trust and trustworthiness environments, the contribution differences between high (low) trust and high (low) trustworthiness environments are negligible and statistically insignificant (Mann-Whitney tests, *p*-values > 0.7). Aligning with the observations from Figure 2, 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:

Result 4: Subjects contribute significantly more in high "trust"/"trustworthiness" environments than in low "trust"/"trustworthiness" environments, with contributions in the random environment falling in between, results that support H3. These outcomes obtain although subjects are provided with information solely about members' relative rankings *in the trust game, without information about behaviors or their distribution. When making decisions in our VCM games, subjects treat relative trust and relative trustworthiness of other group members interchangeably.*[28]

5.4. Conditional cooperation and guilt aversion in the VCM

To delve more deeply into the relationship between first-order beliefs and contributions in the VCM and test our remaining hypotheses, we begin by estimating 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

[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 in four of six 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 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.

[28] Note that to assess the accuracy of subjects' beliefs about others' contributions in each environment, we cannot simply compare Figures 3 and 2, since Figure 3 displays the average contribution made by *all* subjects for their respective groups of each environment, but the sets of four subjects eligible to be in each environment differs from subject to subject. For further analysis and discussion, see Table A.2, Figure A.1, and the associated notes.

themselves). As stated in H6, guilt aversion is a channel through which second-order beliefs might affect contributions, and it is an empirical question which order of beliefs is more prominent in shaping contributing behaviors. Table 3 shows the results of a separate linear regression with robust standard errors clustered by session for each of the five environments, and one in which the contributing behavior of all environments are pooled. In each regression, first-order beliefs obtain highly significant positive coefficients, supporting H3 at the level of individual decisions, 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 is stronger than that between second-order beliefs and contributions.[29]

Table 3. 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***	1.027***	0.859***	1.156***	1.158***	0.982***
	(0.204)	(0.144)	(0.113)	(0.0980)	(0.122)	(0.106)
Second-Order Beliefs	-0.0642	-0.000773	0.176	-0.0513	-0.117	-0.0461
	(0.280)	(0.0906)	(0.142)	(0.107)	(0.196)	(0.120)
Constant	0.119	0.394	-1.407	-0.274	-1.951	0.241
	(0.662)	(0.745)	(0.730)	(0.371)	(1.192)	(0.372)
Observations	120	120	120	120	120	600
R-squared	0.522	0.477	0.448	0.591	0.529	0.580

Standard errors in parentheses are clustered at the session level. *** 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 use a 2-step differencing approach. We first regress second-order beliefs on first-order beliefs, and we save the residuals because they capture the part of second-order beliefs that is orthogonal to first-order beliefs. Then, 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

[29] For completeness, we note that if we include in regressions only first-order beliefs and a constant, or only second-order beliefs and a constant, each belief variable 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.

impact contribution decisions in the VCM. However, the results (shown in Table A.3) are quantitatively and qualitatively similar to those of Table 3: both in estimates for all decisions pooled, and in estimates environment by environment, the only significant coefficient is that of first-order beliefs.

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	0.555***	0.812***	0.653***
	(0.104)	(0.158)	(0.103)
Difference in Second-Order Beliefs	0.0986	-0.124	0.00680
	(0.0621)	(0.150)	(0.0782)
Constant	1.326	0.917	1.215
	(0.933)	(0.855)	(0.844)
Observations	120	120	240
R-squared	0.242	0.323	0.274

Robust standard errors in parentheses are clustered at the session level. *** 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.

As a final check of the role of beliefs in determining VCM contributions, we study whether between-environment differences in individuals' first- or second-order beliefs in high versus low environments can predict the corresponding differences in their contributions. Column (1) of Table 4 shows regressions using differences between the high and the low trust environments, while column (2) shows the corresponding regressions for differences in high vs. low trustworthiness environments. Column (3) shows regressions on the pooled data of all four environments, making use of the apparent similarity of the effects of relative trust and of relative trustworthiness on subjects' first- and second-order beliefs, and on their behaviors. The estimates suggest that differences in first-order beliefs between the high and low member of a pair of environments significantly predict differences in contributions in the high versus the low environment of the pair, in each of the environment pairs (i.e., those for trust and those for trustworthiness) and for both pairs of environments pooled. Consistent with the previous findings, differences in second-order beliefs are not significantly related to differences in contributions in the presence of differences in first-order beliefs. The insignificant estimates on the constant terms suggest that there is no

tendency for VCM contributions to differ systematically between high and low environments, apart from that associated with the differences in first-order beliefs.

Result 5: Subjects' VCM contributions in each environment are a significant positive function of their first-order beliefs, supporting H3, 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. Our data thus provide little evidence of guilt aversion (H6) as a force in its own right.

5.5. Trustworthiness and conditional cooperation

The estimated coefficients on first-order beliefs in Table 4 can be interpreted as measuring the average 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 (about 0.65 units on average, according to col. (3)). Such an average increase would be consistent with all subjects being of relatively homogeneous inclination to be imperfect conditional cooperators, but it is also 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, and free riders who contribute nothing regardless of their belief about others. Although the size and high significance of the coefficients suggests that conditional cooperators of some kind are likely to be quite numerous, past research suggests that inclinations vary considerably within populations, and our predictive framework also anticipates the possibility of subjects for whom $r_c = 0$, who would be expected not to contribute regardless of environment. Therefore, we now investigate the variation of subject types and test whether reciprocity in the trust game can predict heterogeneity of behavioral types within VCM environments, as anticipated by H4.

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 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.[30] This has the potential advantage that decisions are made without the possibly structuring or even

[30] Using the relationship between beliefs and contributions rather than the conditional contribution approach of past literature can be expected to yield rather similar results, judging by the findings of Fischbacher and Gächter (2010), who determine that the relationship between hypothetical contribution of others and own conditional contribution, as displayed in the elicitation task of Fischbacher *et al*, (2001), is also essentially the same as that between belief of others' contribution and own contribution, in the data of their experiment.

leading frame of a table, and with beliefs and contributions inputted on separate screens, but it has 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 operationalize classifications resembling those of the literature using slightly different criteria. We start by identifying 5 categories: (1) free riders (who never contribute), (2) weak conditional cooperators (whose contributions tend to increase when others' do, but often contributing less than the latter), (3) strong conditional cooperator (similar but usually contributing as much or more than others), (4) altruist (who contribute their full endowment regardless of others' contributions), and (5) other.[31] 13 subjects cannot be assigned to one of the first four categories and are thus classified in category (5), which we exclude from our initial exercises due to its high internal diversity and lack of compelling placement relative to categories (1) to (4). We also consider an expanded classification scheme that permits all except two subjects to be assigned to one or another of seven categories, which add to the above-listed five types "broad free riders" (sometimes contributing 1 and at most once contributing between 2 and 5), "hump-shaped contributors" (whose contributions first rise and then fall as belief rises, inserted after "broad free riders" but before "weak conditional cooperators"), and "broad conditional cooperators (who contribute if others do, but without clear rising trend)."[32] Both variants of our classification exercise show our modal subject to be conditionally cooperative, as in other studies.

Having classified most subjects in terms of a reciprocity type in the VCM, we can now test H4, which captures a core assumption of our analytical framework, namely that each individual can be characterized by a reciprocity parameter r_i that predicts both her trust game returning decisions and her cooperation conditional on her first-order beliefs in the VCM. With both sets of categorical values, we estimate ordered probit regressions to check whether behavioral disposition in the VCM can be predicted by 1st or 2nd mover choices in the trust game. The results are shown in Table 5, with columns (1) – (3) using the initial 4 types covering 107 subjects, and columns (4) – (6) using the 7 types covering 118 subjects. In columns (1) and (2), we find that both sending and returning behavior in the trust game are strongly correlated with types in the VCM ordered from free rider to weak and then strong conditional cooperator, and finally altruist, when either variable is entered. The significance level is slightly stronger for returning (trustworthiness, column (2)), and returning alone is highly significant when both variables are entered in column

[31] More precisely, classification of types (2) and (3) begins by graphing subjects' contributions against their first-order beliefs in each environment. If a subject contributes more than she believes others contribute at least four of five times and her (belief, contribution) pairings are best fit by an upward sloping line, she is a strong conditional cooperator. If she contributes less than she believes others contribute at least twice but the slope of that best fit line is positive, she is a weak conditional cooperator. Because of limited data points, we impose no requirement on the significance of the partial correlation. Additional details are provided in Appendix 4.

[32] We identify the new type "broad conditional cooperator" in order to place among the other types seven subjects who contribute when they believe others do so, but for whom there is either no unique best fit line (because they have the same belief and contribution in all five environments) or the best fit line is strictly vertical or horizontal (in the latter case, lying below full contribution, thus also not fitting the altruist type). See Appendix 4 for details and scatter plots between contributions and first-order beliefs, shown aggregated for each type, and for each individual subject.

(3). This stronger impact of trustworthiness is predicted by our explanatory framework and confirms H4. 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.[33] The

Table 5. Contribution type (e.g. conditional cooperation) predicted by trust and trustworthiness

Dependent Variable：	4-type specification			7-type specification		
VCM Types, Ordered	(1)	(2)	(3)	(4)	(5)	(6)
Sending Percentage (trust)	1.427**		0.958	1.438***		1.077
	(0.583)		(0.647)	(0.554)		(0.677)
Average Returning Percentage (trustworthiness)		1.766***	1.161***		1.557***	0.862*
		(0.326)	(0.344)		(0.346)	(0.475)
Cut points						
Observations	107	107	107	118	118	118
Pseudo R-Squared	0.065	0.064	0.085	0.053	0.040	0.062
p-value of Wald test	-	-	0.001	-	-	0.001

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=broad conditional cooperator, 6=strong conditional cooperator, and 7=Altruist. For estimated cut point values and standard errors, see Appendix Table A. (cont.) Standard errors that are clustered at the session level 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.

estimates in columns (4) – (6) are similar, with both trusting (column 4) and trustworthiness (column 5) individually predicting reciprocity type in the VCM, the latter somewhat more strongly, and with the latter variable only showing some marginal significance in column (6). A Wald test confirms that sending and returning jointly predict cooperation tendency or type, in that column, with *p*-value 0.001.[34] Therefore, we end this discussion with the following conclusion:

Result 6: The relationship between subjects' first-order beliefs and their contributions by environment permits classifications suggesting that the majority of our subjects are conditional cooperators, with a small share of free riders and a few altruists and hump-shaped contributors. Subjects' behaviors in the trust game are a good predictor of their type in the VCM at the

[33] Regarding the marginal effects, estimates (1) – (3) imply that a 1% increase in sending and returned amounts are associated with a 0.08% and 0.34% increase on the probability of being an altruist and a strong conditional cooperator at the 1% significance level, respectively.

[34] While neither variable is individually significant, the coefficient on trustworthiness comes much closer to marginal significance than that on trust.

individual level. In particular, the greater the proportion returned (and with slightly less predictive power, the more sent as first-mover) in the trust game, the less likely to be a free rider, with likelihood of being a weak conditional cooperator, a strong conditional cooperator, and an altruist increasing in turn as these trust game choices rise, supporting H4.

The results thus far described leave untested only H5, which predicts that more (less) trusting and trustworthy subjects will contribute more (less) in given VCM environments. Several of the results already reported combine to yield a strong expectation that this must be true. First, we confirmed that trusting and trustworthiness themselves are highly correlated (H1). Second, we found that beliefs that others in an environment are relatively (un)trustworthy or (un)trusting lead to higher (lower) first order beliefs about those others' contributions in the VCM, implying variation in expected average contributions along the lines illustrated in Figure 2 (H3). Third, we found that subjects are on average conditionally cooperative, with greater likelihood of conditional cooperation among more trusting and trustworthy subjects (H4). Fourth, we found that for given VCM environments, more (less) trustworthy and trusting subjects have more (less) optimistic beliefs about what others will contribute on average (H2b). All of these results add up to the expectation that H5 will hold. That expectation is confirmed by regressions, which are displayed in the Appendix (Table A.5) to conserve space.

Result 7: More (less) trustworthy and more (less) trusting subjects contribute more in given VCM environments and overall, confirming H5.

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 our subjects' behaviors are not highly atypical, 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 being probable 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 and trustworthiness. Ours is the first investigation of how information about *relative* 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 their VCM environments, 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 regarding higher ranked trustors and higher ranked trustees were indistinguishable, as were contribution expectations regarding 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 is quite trusting and trustworthy in the trust game and that these same individuals are quite cooperative in the VCM. They believe another sizeable fraction is 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 the assumptions 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 raising the productivity of resources, the line connecting trust with economic efficiency and growth appears likely be supported by the disposition of reciprocity and the variation of beliefs about its strength in different settings. While the potential for such a disposition to be present in human populations may well have deep roots in human evolution (Wilson, 2004), 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.

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. Kzagel 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 Eeview* (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.

Sobel, Joel, "Interdependent Preferences and Reciprocity," *Journal of Economic Literature* (2005) 43: 392 – 436.

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.

Wilson, Edward O., *On Human Nature, 2nd Edition*. Cambridge: Harvard University Press (2004).

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 ONLY

Appendix to Kim, Putterman, and Zhang, 2019

"Trust, Beliefs, and Cooperation: An Experiment"

Overview

Appendix 1: Instructions

Appendix 2: Screenshots of the five environments and end-of-session survey

Appendix 3: Additional Tables and Figures

Appendix 1: 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*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.

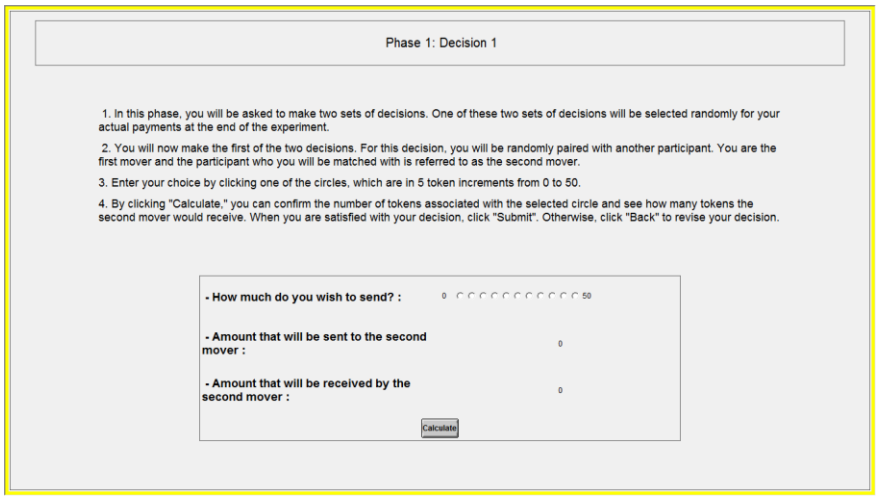
Each of you will be making decisions first as a first mover, and then as a second mover. Only one of these roles can be selected for payment, and for whichever role that is, the computer will randomly select a participant to pair you with, so you will never play the second mover part against yourself 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.

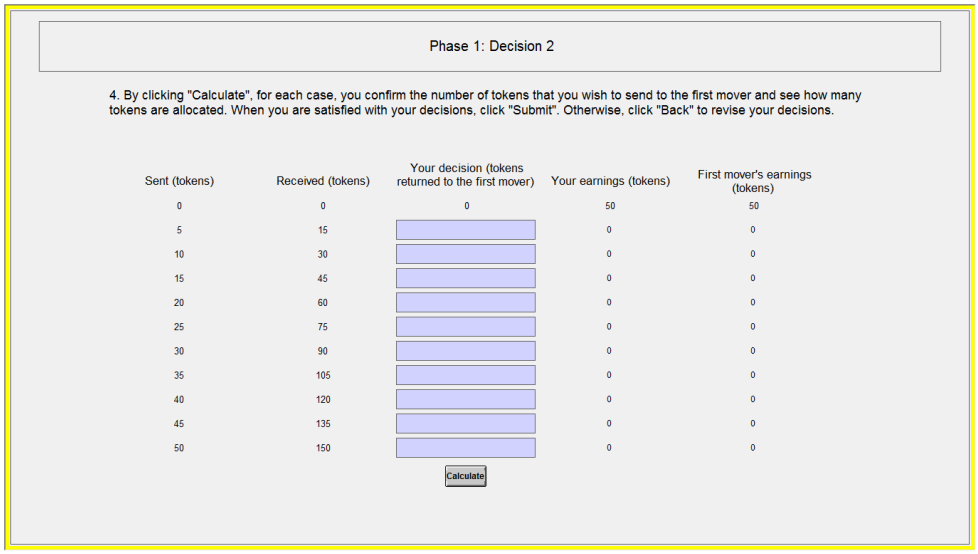
Your decision is then final.



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.**



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 (1^{st}) to the highest (20^{th}) 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 1^{st} movers, while in Environment 3 you are grouped with others who on average sent relatively large amounts as 1^{st} 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 (1^{st}) to the highest (20^{th}) 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**.

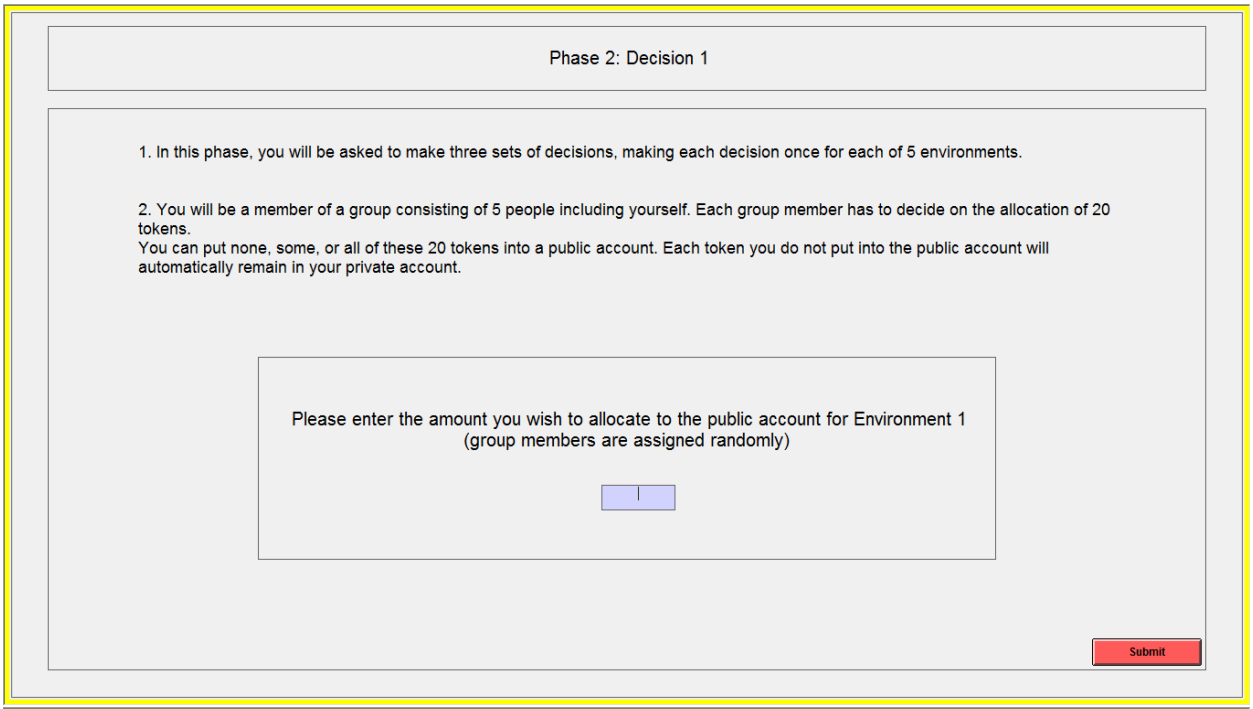
Example:

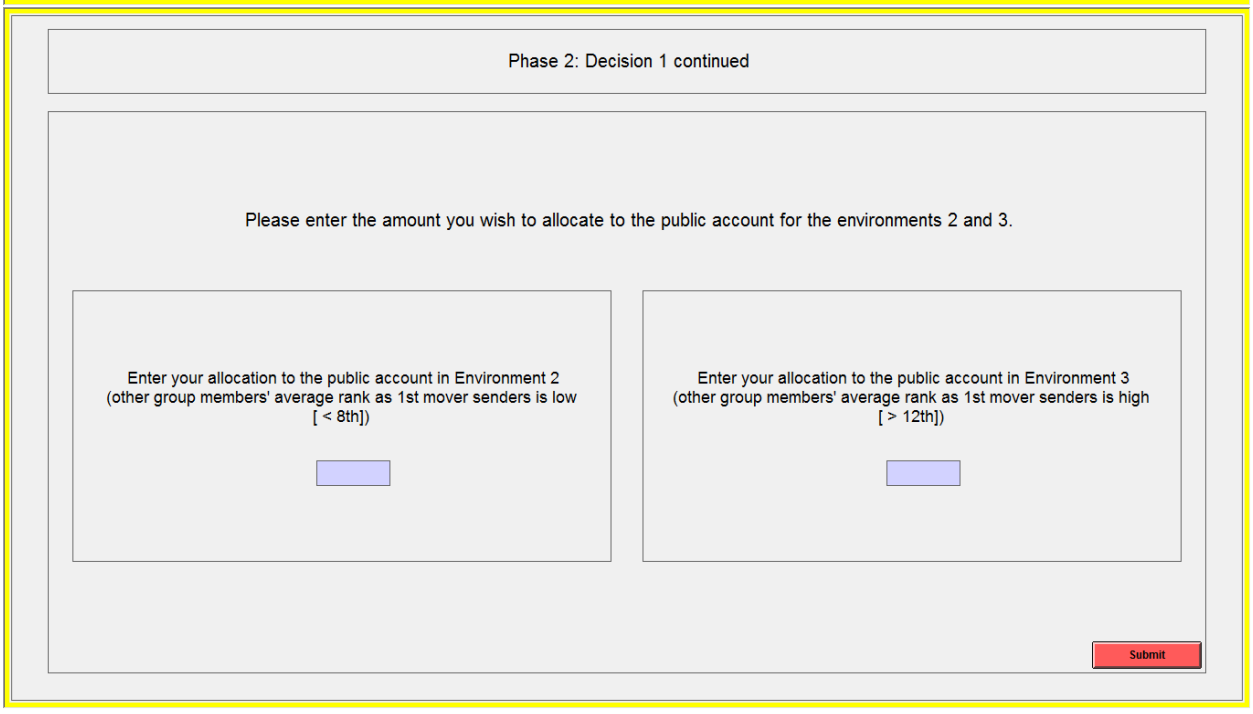
Suppose Environment 1 is chosen for payment, and your estimate in Decision 2 is in the correct range but that in Decision 3 is not; moreover, you contributed 10 tokens in Decision 1 and the public account ends up with 45 tokens. In the end, you would get: {(20-10) + 0.4*45} + 5 + 0 = 33 tokens, which 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 2: Screenshots of the five environments and end-of-session survey





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

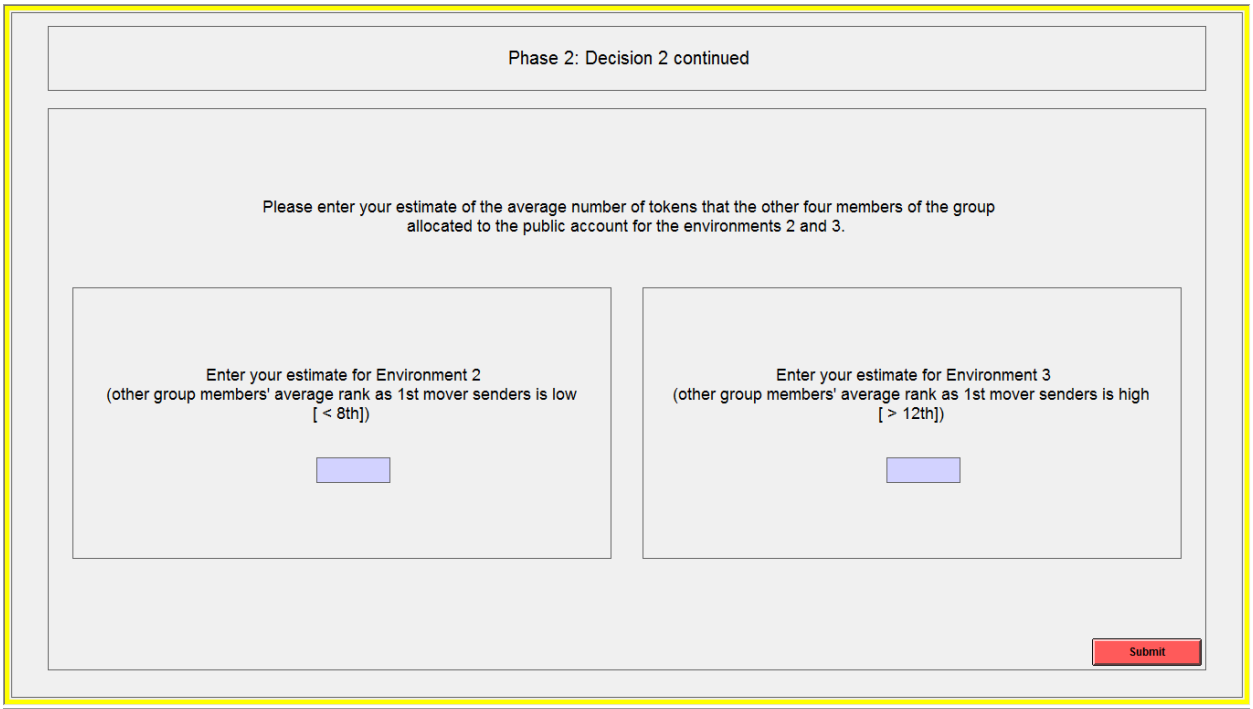
Phase 2: Decision 2

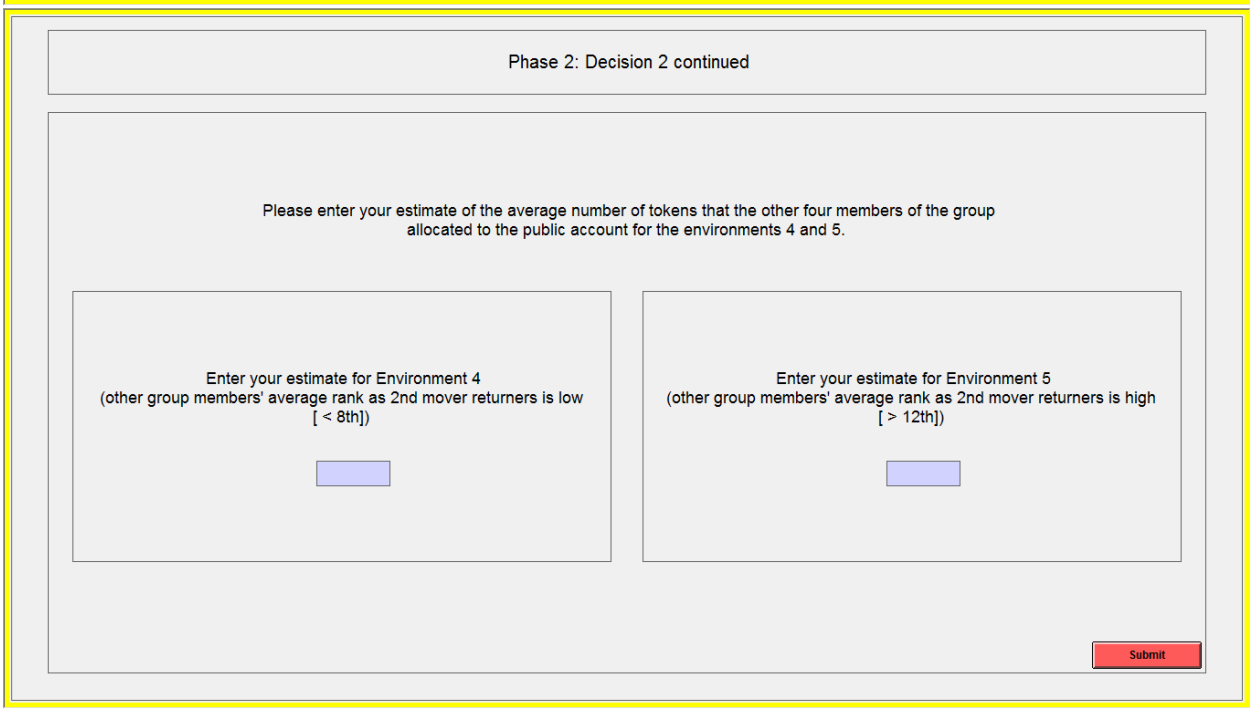
1. In the previous decision, you and the other four members in each of five possible groups decided how many tokens they allocated to the public account.

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

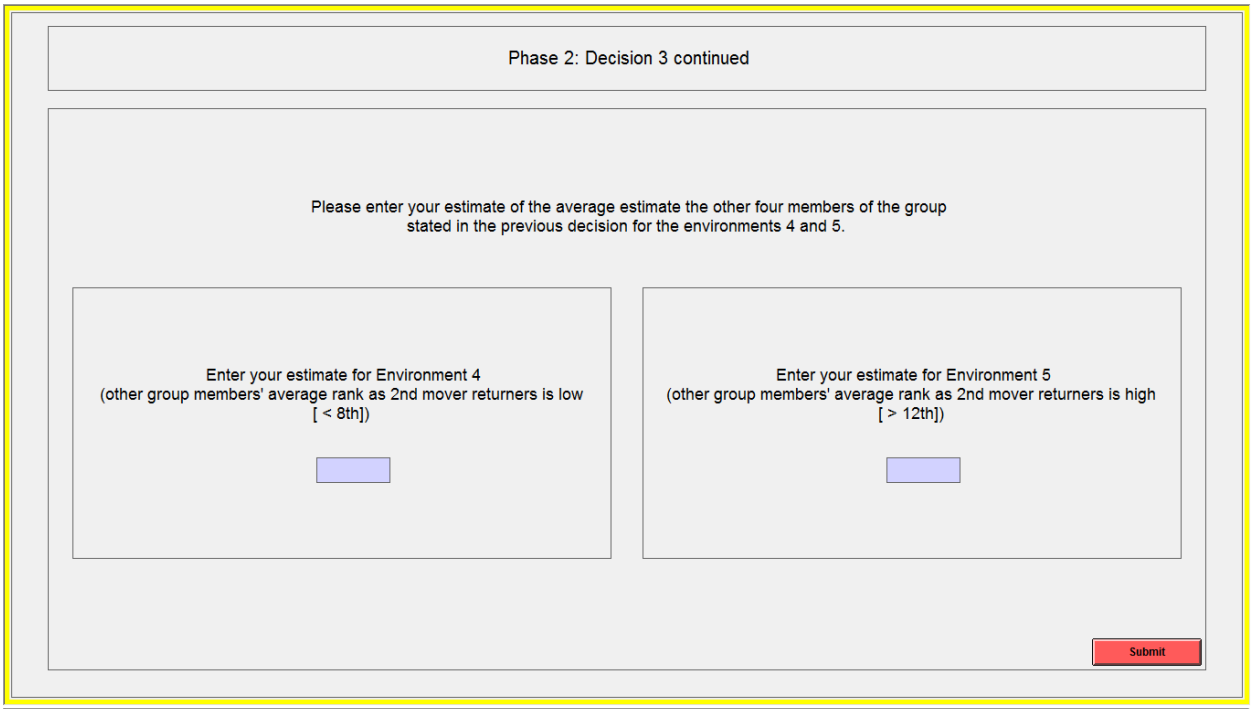
Phase 2: Decision 3 continued

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



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

Appendix 3: Additional Tables and Figures

Table A.1 Trust game behaviors and first-order beliefs to 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	2.17	3.77***	2.72	3.26***	3.42	3.96**
in Trust Game	(1.558)	(0.869)	(1.412)	(0.535)	(1.868)	(1.046)
Constant	7.47***	2.92***	11.44***	3.16***	11.3***	6.91***
	(0.736)	(0.349)	(0.783)	(0.315)	(0.979)	(0.553)
Observations	120	120	120	120	120	600
R-squared	0.019	0.09	0.029	0.069	0.048	0.048

(a) Univariate model with trust as explanatory variable

Dependent Variable:	(1)	(2)	(3)	(4)	(5)	(6)
First-order beliefs	Random	Low Trust	High Trust	Low Trustworthiness	High Trustworthiness	Pooled and Differenced
Average Returning %	3.22***	4.54***	4.97**	3.78***	5.54***	5.26***
in Trust Game	(0.661)	(0.9)	(1.431)	(0.615)	(1.171)	(0.835)
Constant	7.30***	2.96***	10.94***	3.24***	10.89***	6.8***
	(0.392)	(0.318)	(0.607)	(0.384)	(0.627)	(0.427)
Observations	120	120	120	120	120	600
R-squared	0.028	0.087	0.064	0.062	0.085	0.057

(b) Univariate model with reciprocity as explanatory variable

Standard errors in parentheses are clustered at the session level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The linear regression models in panels (a) and (b) correspond to that of Table 2 which regresses first-order beliefs of contributions to the VCM on both sending and returning amount in the Trust game.

Table A.2 First-order beliefs and actual contributions by environment and own trustworthiness rank.

Environment	Low Trust				High Trust			
Criterion	average trust rank ≤ 8				average trust rank ≥ 12			
(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)
rank	avg. belief	avg. contr.	(ii) ÷ (iii)	# elig. groups	avg. belief	avg. contr.	(vi) ÷ (vii)	# elig. groups
1	4.8	3.7	1.31	774	10.2	15.3	0.67	774
2	4.5	3.6	1.25	663	12.3	15.2	0.81	965
3	2.2	3.1	0.69	632	13.3	15.1	0.88	1010
4	1.8	3.4	0.54	1179	8.0	15.0	0.53	371
5	1.0	3.3	0.30	782	12.7	14.8	0.85	826
6	2.8	3.0	0.93	969	12.5	14.7	0.85	554
7	5.2	2.9	1.80	844	9.5	14.6	0.65	711
8	6.7	3.1	2.12	507	11.5	14.5	0.80	1191
9	3.8	3.0	1.26	815	14.2	14.2	1.00	736
10	4.0	2.9	1.40	389	14.5	14.4	1.01	1400
11	3.0	2.9	1.05	728	10.3	14.3	0.72	853
12	4.2	2.8	1.48	609	11.3	14.0	0.81	968
13	6.8	2.7	2.54	519	14.3	13.7	1.05	1177
14	6.2	2.7	2.32	318	13.3	14.1	0.94	1436
15	3.2	2.6	1.24	538	14.5	13.3	1.09	1123
16	4.2	2.5	1.65	397	11.2	13.8	0.81	1266
17	3.5	2.5	1.43	463	14.3	13.1	1.09	1202
18	5.3	2.4	2.24	304	13.8	13.3	1.04	1449
19	5.3	2.2	2.38	191	10.7	13.5	0.79	1736
20	9.3	2.1	4.40	225	17.7	12.8	1.38	1694

(a) Environments determined by trust, ranked by trustworthiness

Environment	Low Trustworthiness				High Trustworthiness			
Criterion	average trustworthiness rank ≤ 8				average trustworthiness rank ≥ 12			
(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)
rank	avg. belief	avg. contr.	(ii) ÷ (iii)	# elig. groups	avg. belief	avg. contr.	(vi) ÷ (vii)	# elig. groups
1	4.2	3.7	1.14	508	9.8	14.4	0.68	44
2	4.8	3.5	1.39	508	12.5	14.3	0.87	61
3	2.2	3.9	0.56	482	12.5	14.2	0.88	84
4	1.8	3.6	0.51	443	7.8	14.1	0.56	111
5	1.3	3.9	0.34	398	13.0	14.0	0.93	144
6	3.7	3.7	0.99	353	13.7	13.8	0.99	184
7	5.0	3.5	1.45	311	9.3	13.8	0.68	226
8	6.0	3.4	1.79	269	11.5	13.7	0.84	271
9	3.8	3.4	1.12	231	14.0	13.6	1.03	320
10	4.3	3.3	1.31	195	15.2	13.3	1.14	371
11	3.7	3.3	1.10	166	11.2	13.4	0.83	430
12	4.2	3.3	1.27	136	12.3	13.4	0.92	491
13	7.0	3.2	2.18	109	13.5	13.1	1.03	555
14	7.2	3.2	2.25	84	12.3	12.8	0.96	619
15	3.0	3.2	0.95	61	14.5	12.8	1.14	687
16	4.8	3.1	1.55	44	11.3	13.0	0.87	758
17	2.8	3.1	0.93	31	13.7	12.6	1.08	823
18	5.5	3.1	1.80	20	13.2	12.7	1.04	878
19	3.8	3.0	1.27	13	13.8	12.7	1.09	909
20	9.5	2.9	3.25	8	17.5	12.3	1.42	909

(b) Environments determined by trustworthiness, ranked by trustworthiness

Notes: For panels (a) and (b): Explanation of variables: "rank" is subject i's rank for amount returned in the Trust Game when subject j sends the maximum amount (50 tokens). Ties were broken randomly when multiple subjects returned the same amount within a session. "avg. belief" is the average first order belief about other group members' contribution to the group account by the groups of four others, who, with subject i, meet the environment's criterion for average rank; "avg. contr." is the average actual contribution to the group account by all possible groups of four who, with subject i, meet the environment's criterion for average rank; "(ii) ÷ (iii)" is the ratio of variable (ii) to variable (iii); "# elig. groups" is the number of eligible sets of four subjects who could be grouped with each subject and fulfill the stipulation about average trust or trustworthiness rank, calculated within each session and then averaged across sessions. Note that the average contribution of others used to complete variable (iii) is calculated for each subject over all eligible groups in his or her session.

(c) Average contribution belief, overall contribution, and contribution by those eligible to constitute a group, by environment

Environment	Average 1st order belief about others' contribution	Average contribution of those others eligible	Average contribution decision of all others
Random	8.31	8.59	8.59
Low trust	4.39	2.87	4.90
High trust	12.51	14.18	11.48
Low trustworthiness	4.43	3.36	4.61
High trustworthiness	12.63	13.39	11.22

Notes: Average contribution of those eligible to be in the environment with the subject forming the belief is calculated as described for the panels above. Average contribution decision of all others is the average contribution decision for the environment in question of all nineteen subjects in the belief-former's session.

Dependent Variable:	(1)	(2)	(3)	(4)	(5)
Belief to Actual Ratio of Others' Contributions	Low Trust	High Trust	Low Trustworthiness	High Trustworthiness	Pooled
Trustworthiness Rank	0.10***	0.02***	0.06**	0.02***	0.05***
	(0.034)	(0.007)	(0.027)	(0.006)	(0.014)
Constant	0.54*	0.68***	0.70**	0.71***	0.66***
	(0.311)	(0.069)	(0.255)	(0.068)	(0.118)
Obsv.	20	20	20	20	80
R-squared	0.687	0.154	0.576	0.141	0.220

(d) Return amounts in Trust Game and belief of others' contributions in the VCM

Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. When ranking subjects by their returning (trustworthiness) amounts, ties were broken by random number generation. The dependent variable, "belief to actual ratio," is the number in the third column from left of each portion of panels (a) and (b) above (for example, columns (iv) and (viii)), while the independent variable rank corresponds to the row headings of those panels (i.e., column (i) in each case).

Appendix Table A.2 general note. These tables address the relationship between subjects' beliefs about others' contributions in the four VCM environments that feature group formation based on relative rank of choices in the trust game, and the actual average contribution of any four subjects who were eligible to be in the relevant environment with a subject of given trustworthiness rank. Panels (a) and (b) feature a row for the subjects of each trustworthiness rank, and for each of the four environments (low and high trust environments, panel (a), low and high trustworthiness environments, panel (b)) the average belief about others' contributions by those of this rank, the average actual average contribution of sets of four subjects eligible for the group, the ratio of belief to actual, and the average number of sets of four eligible to have been grouped with the individuals of indicated rank. That is, rather than calculate the average contribution of those actually assigned to be grouped with a subject *i* or rank *q* for trustworthiness, we identified all possible groups of four that met the criterion, and calculated the average contributions across those groups of four. Calculations, including determining the number of eligible sets of four, were performed for each of our six sessions, the results were then averaged. By looking at the numbers eligible columns, it can be confirmed that, for example, an individual ranked 20^{th} (1^{st}) for trustworthiness would have fewer (more) potential sets of four with whom to be grouped while meeting the grouping criteria. The main finding is that, consistent with introspection bias, the most trustworthy subjects tended to overestimate the contributions of others in each environment by more than did the least trustworthy, even though a rational consideration of the impact of own trustworthiness on who could be grouped with oneself to meet the criteria meant, for example, that a more trustworthy individual could be grouped with others who were on average somewhat less trustworthy than a less trustworthy individual, when meeting the criteria for high trustworthiness group. The regressions of panel (c) confirm that overestimation of other group members' contributions is significantly increasing in own trustworthiness, consistent with introspection bias and H2.b.

Table A.3 Contribution and Beliefs in the VCM, 2-step verification

(a) Pooling all treatments, both stages displayed

Dependent Variable:	1st stage Second Order Belief	2nd stage Contribution
First-Order Belief	0.809***	0.945***
	(0.0246)	(0.0568)
Residual from the first stage		-0.0461
		(0.120)
Constant	1.588***	0.168
	(0.255)	(0.383)
Observations	600	600
R-squared	0.736	0.580

Standard errors in parentheses are clustered at the session level. *** p<0.01, ** p<0.05, * p<0.1.

(b) Environment by environment, displaying 2nd stage only

Dependent Variable: Contributions	(1) Random	(2) Low Trust	(3) High Trust	(4) Low Trustworthiness	(5) High Trustworthiness
First-Order Beliefs	1.032***	1.026***	0.970***	1.117***	1.070***
	(0.0717)	(0.120)	(0.0570)	(0.0880)	(0.0480)
Residuals	-0.0642	-0.000773	0.176	-0.0513	-0.117
	(0.280)	(0.0906)	(0.142)	(0.107)	(0.196)
Constant	0.0113	0.393	-0.644	-0.344	-2.306**
	(0.608)	(0.784)	(0.705)	(0.358)	(0.762)
Observations	120	120	120	120	120
R-squared	0.522	0.477	0.448	0.591	0.529

Second stage regressions; residuals calculated in first stage as in Table 4 of paper. Standard errors in parentheses are clustered at the session level. *** p<0.01, ** p<0.05, * p<0.1

Table A.4 Estimated cut points from Table 5

Dependent Variable：	4-type specification			7-type specification		
VCM Types, Ordered	(1)	(2)	(3)	(4)	(5)	(6)
Cut point 1	-0.612***	-0.639***	-0.479***	-0.688***	-0.730***	-0.577***
	(0.156)	(0.123)	(0.144)	(0.159)	(0.136)	(0.139)
Cut point 2	0.857***	0.843***	1.051***	-0.563***	-0.602***	-0.447***
	(0.24)	(0.105)	(0.247)	(0.154)	(0.09)	(0.128)
Cut point 3	2.292***	2.255***	2.497***	-0.450***	-0.492***	-0.332**
	(0.455)	(0.293)	(0.448)	(0.157)	(0.0937)	(0.131)
Cut point 4				0.785***	0.704***	0.921***
				(0.241)	(0.103)	(0.213)
Cut point 5				0.957***	0.872***	1.096***
				(0.234)	(0.0918)	(0.204)
Cut point 6				2.339***	2.219***	2.484***
				(0.443)	(0.266)	(0.405)

Note: see notes to Table 5 in text. The estimates above were omitted from that table to save space.

Table A.5 Behavior in the Trust game and contributions in the VCM

Dependent Variable:	(1)	(2)	(3)	(4)	(5)	(6)
Contributions	Random	Low Trust	High Trust	Low Trustworthiness	High Trustworthiness	Pooled
Sending Percentage	6.398	6.207*	7.406*	2.966*	6.468	5.889*
	(3.266)	(2.993)	(3.470)	(1.451)	(3.668)	(2.703)
Average Returning Percentage	5.559	1.341	7.217	5.033**	7.622	5.354*
	(3.082)	(2.210)	(3.975)	(1.476)	(3.808)	(2.408)
Constant	4.352***	2.050	6.331***	1.877*	6.306***	4.183***
	(0.871)	(1.100)	(0.819)	(0.931)	(0.719)	(0.698)
Observations	120	120	120	120	120	600
R-squared	0.181	0.135	0.245	0.117	0.233	0.148
p-value for the Wald test	0.0008	0.0375	0.0005	0.0413	0.0004	0.0002

Standard errors in parentheses are clustered at the session level.
*** p<0.01, ** p<0.05, * p<0.1

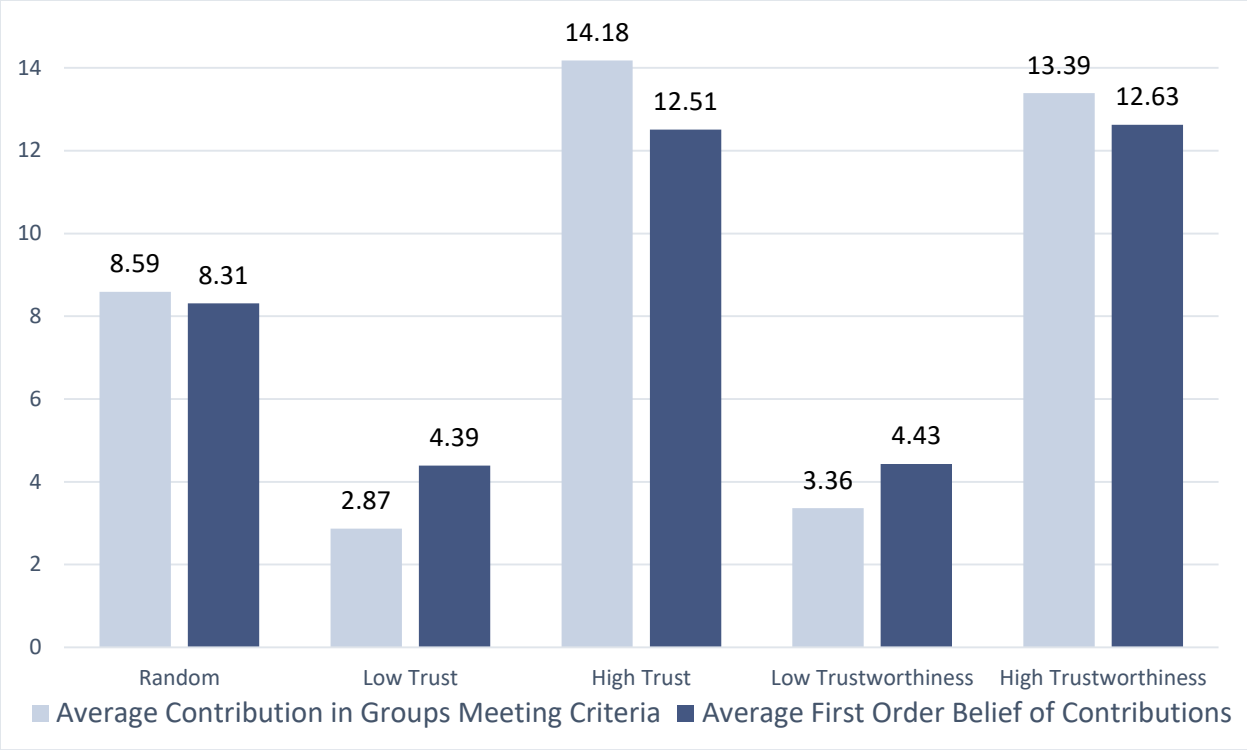
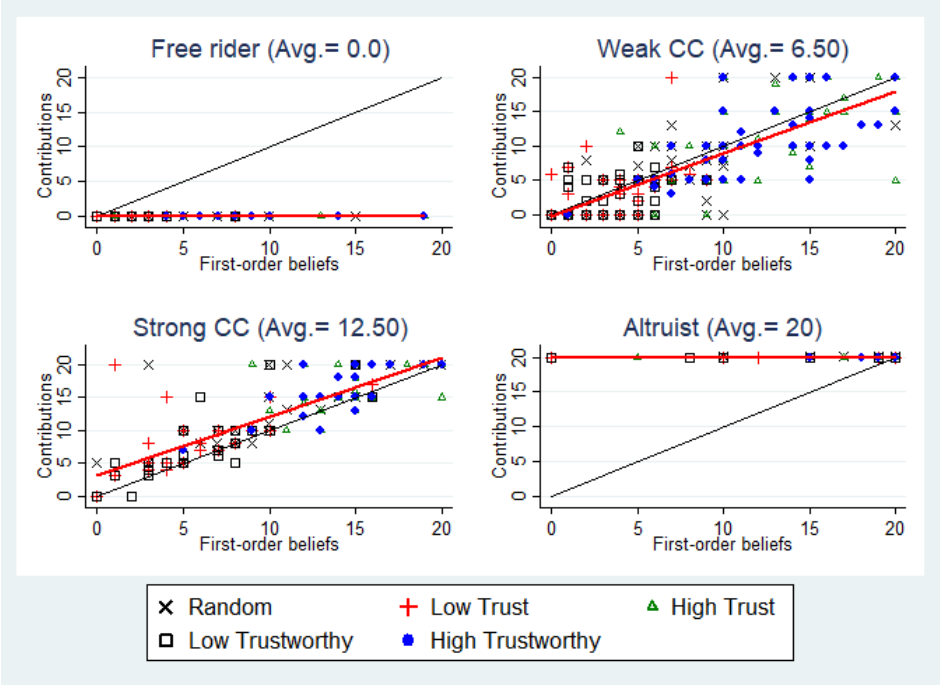


Figure A.1 Beliefs and Actual Contributions to the VCM Environments
(averaging the rows of Table A.3 panels (a) and (b), plus corresponding data for Random environment)

Further note on beliefs, general average contributions by environment, and average contributions of eligible subjects by environment. In our exercises comparing beliefs with actual contributions by environment, prior to Result 3, we used averages of eligible sets of group members, only. Comparing average contributions of actually eligible others by environment to average contribution of all subjects by environment, shown in Figure 3, finds them to be identical for the Random environment, but an average of around 21% higher than Figure 3's levels for the High Trust and High Trustworthiness environments and about 34% lower than Figure 3's levels for the Low Trust and Low Trustworthiness environments. The average first order belief about others' contributions turns out to be higher than the actual average contribution of the others eligible for the low trust and low trustworthiness environments, by 1 or 2 tokens, and lower than the actual average contribution of the eligible others for the random, high trustworthiness, and high trust environments, by less than 1token in the first two cases and nearer to 2 tokens in the last. The anticipated average contribution (belief) lies between (a) the average contribution of all subjects (without consideration of eligibility) in each environment, and (b) the average contribution of those eligible to be grouped with the estimator, given her own trust or trustworthiness rank. However, belief is closer to the overall average than to the average for eligible group members in four out of five environments (see Table A.2 panel (c)). Overall, then, subjects were rather successful in anticipating how others would respond to the different environments, in general, but not as successful in taking the rank-based criteria for determining group membership in each environment into account when forming their beliefs.

Note on definitions of contribution types in 5-type and 8-type classification schemes.

1. Five type scheme

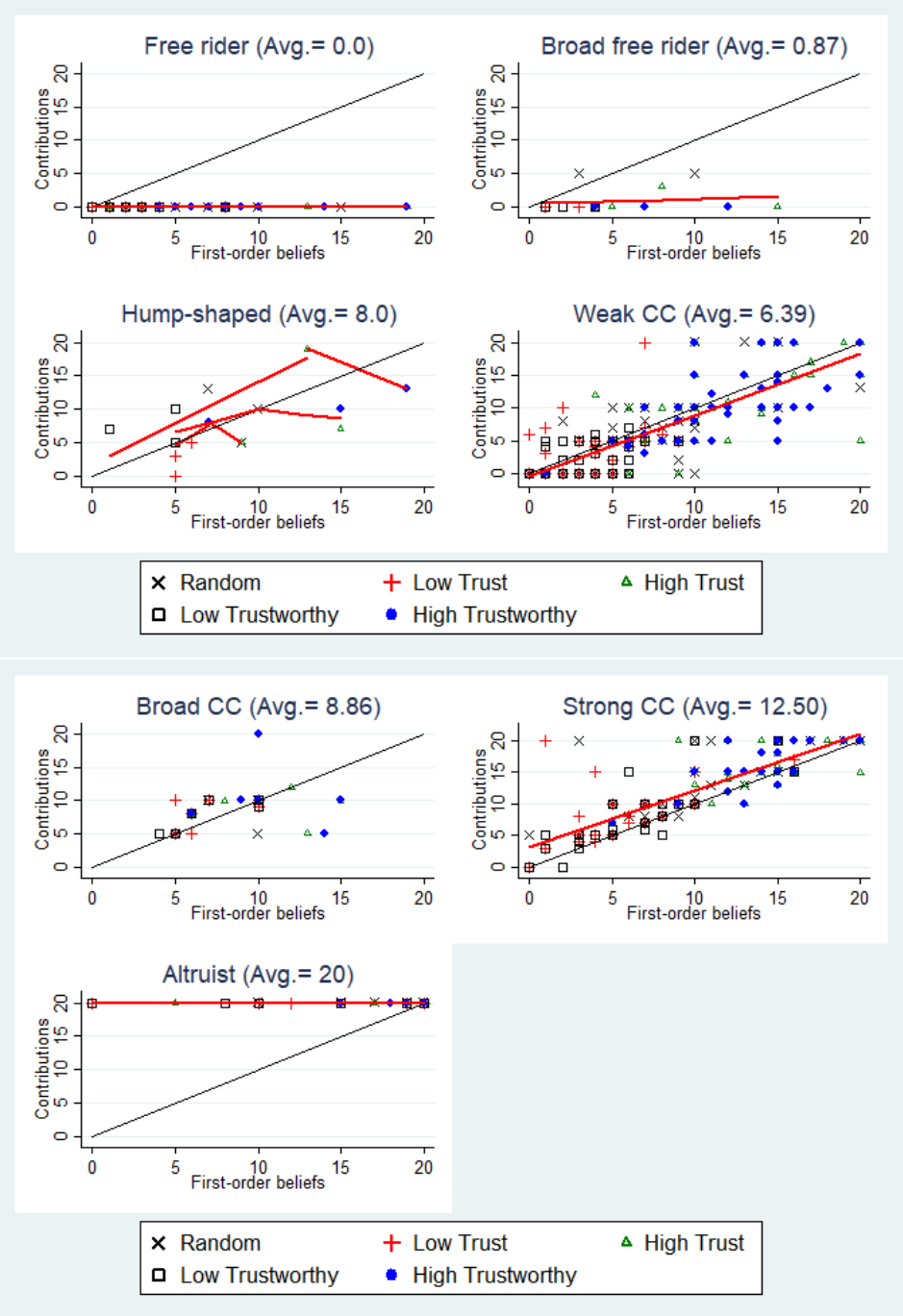


Note: The best fitted lines at the aggregated level and the 45-degree lines in red and black, respectively.

Figure A.2 Contribution and First-Order Beliefs in the VCM by 5 Types

Type 1: Free rider – contributes 0 in all environments.
Type 2: Weak conditional cooperator – contributes more than 1 in at least 2 environments **and** best fit line is upward sloping **and** at least 2 contributions are below the belief about what others contribute in the environment. Note that in all cases that require an upward sloping best fit line, we impose no requirement of minimum correlation level because the limitation of having at most five distinct points and in many cases fewer such points makes a correlation threshold impractical in comparison with studies such as Thöni and Volk (2018) that always have twenty-one points to fit.
Type 3: Strong conditional cooperator – in at least four of the five environments, the subject contributes as much as or more than what she believes others contribute **and** she contributes positive amounts in at least 2 environments **and** the best fit line is upward sloping.
Type 4: Altruist – contributes 20 in every environment.
Type 5: Other – cannot be assigned to any of the four types above.

2. Eight type scheme



Note: The best fitted lines at the aggregated level and the 45-degree lines in red and black, respectively. For Hump-shaped contributors, we draw two-segment fitted lines for each individual. For Broad conditional cooperators, there is no fitted line as some subjects do not have a positive slope.

Figure A.3 Contribution and First-Order Beliefs in the VCM by 8 Types

Types 1. Free rider, 4. weak conditional cooperator, 6. strong conditional cooperator and 7. altruist are defined as above, except that if the definition of "hump shaped" applies, an individual can be reclassified out of the conditional cooperator type.[35]

Type 2: Broad free rider – contributes 0 or 1 in four or more environments and never contributes more than 5.

Type 3: Hump-shaped contributor – contributes a positive amount above 1 in at least two environments and displays a minimum of three distinct (belief, contribution) pairings, q, r and s, such that (i) belief at q < belief at r < belief at s, (ii) contribution at q < contribution at r > contribution at s, (iii) the best-fit line for r and points to its left is upward sloping, and the best-fit line for r and points to its right is downward sloping, and (iv) if one defines the line segments corresponding to the best-fit lines as the segments within the observed range of the subject's beliefs about others' contributions in the five environments, each ending at the horizontal coordinate of r, then the lengths of the two line segments are not too unequal, specifically neither segment exceeds the other in length by a factor greater than three. The reason for requirement (iv) is our judgment that a very small downturn in the neighborhood of the highest observed belief level is insufficient to remove a subject from one of the conditional cooperator categories.

Type 5: Broad conditional cooperator – contributes a positive amount in at least two environments **and** contributes more than 5 in at least one environment **and** either of the following holds: (i) all five beliefs and contributions coincide, or (ii) there are at least two different (belief, contribution) pairs, and the best fit line joining them is parallel to the vertical or horizontal axis.[36]

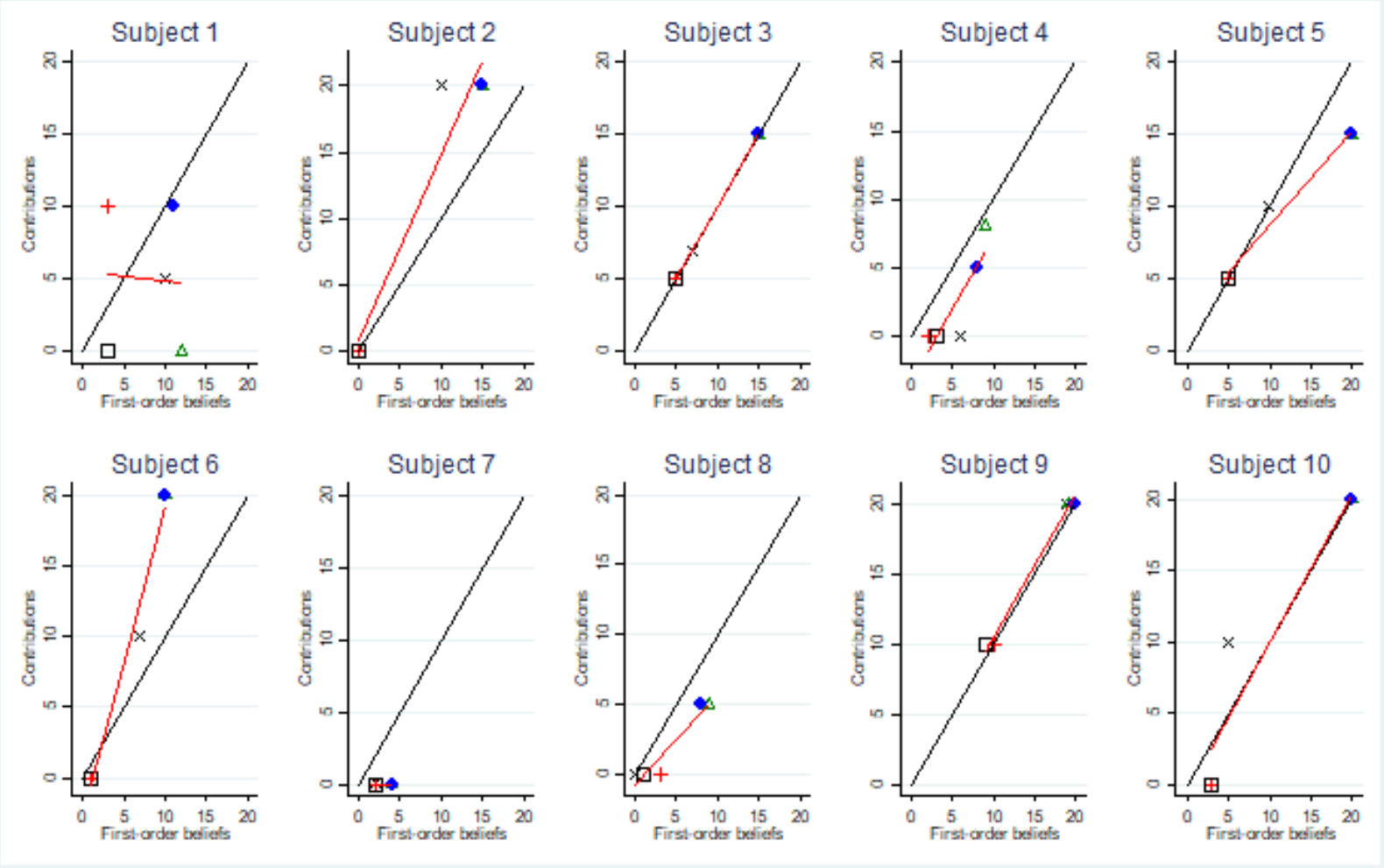
[35] That is, rather than formally amending previous type descriptions, we simply stipulate that the "hump-shaped" category takes priority whenever its definition is met. We have two cases that are reclassified from weak conditional cooperator to hump-shaped when we use this scheme, and only one other hump-shaped case, which is a case reclassified from the "other" category of the five type scheme.

[36] If classification were based on distinct contribution choices for 21 possible average contributions by others, the single point and perfectly vertical case would be ruled out, and it would also then be counter-intuitive to call an unchanging contribution level "conditional cooperation." However, the cases classified by us as "broad conditional cooperator" entail one or a few (belief, contribution) pairs with little distance between them. Although their decision patterns do not (strictly speaking) demonstrate that what these individuals are willing to contribute is an increasing function of what they believe others to contribute, we think it reasonable to see them as showing considerable willingness to contribute if they believe others will do so. Failure of an upward sloping best fit line to obtain may have resulted from lack of variation of belief about what others will contribute (2 cases), or vertically or horizontally adjacent points may appear due to absence of a table format in which the subject can check for a consistent pattern. One also cannot rule out the possibility of influences on contribution other than the belief about others' contributions—for example, some contributions could bear a vertical relationship because a subject might wish to contribute more when dealing with relatively trustworthy counterparts than with the relatively trusting ones, even if both are expected to contribute the same amount on average, due to different judgments of their deservingness. Finally, we were prepared to designate a horizontal pattern at close to the full endowment, say 18 or 19 points, as

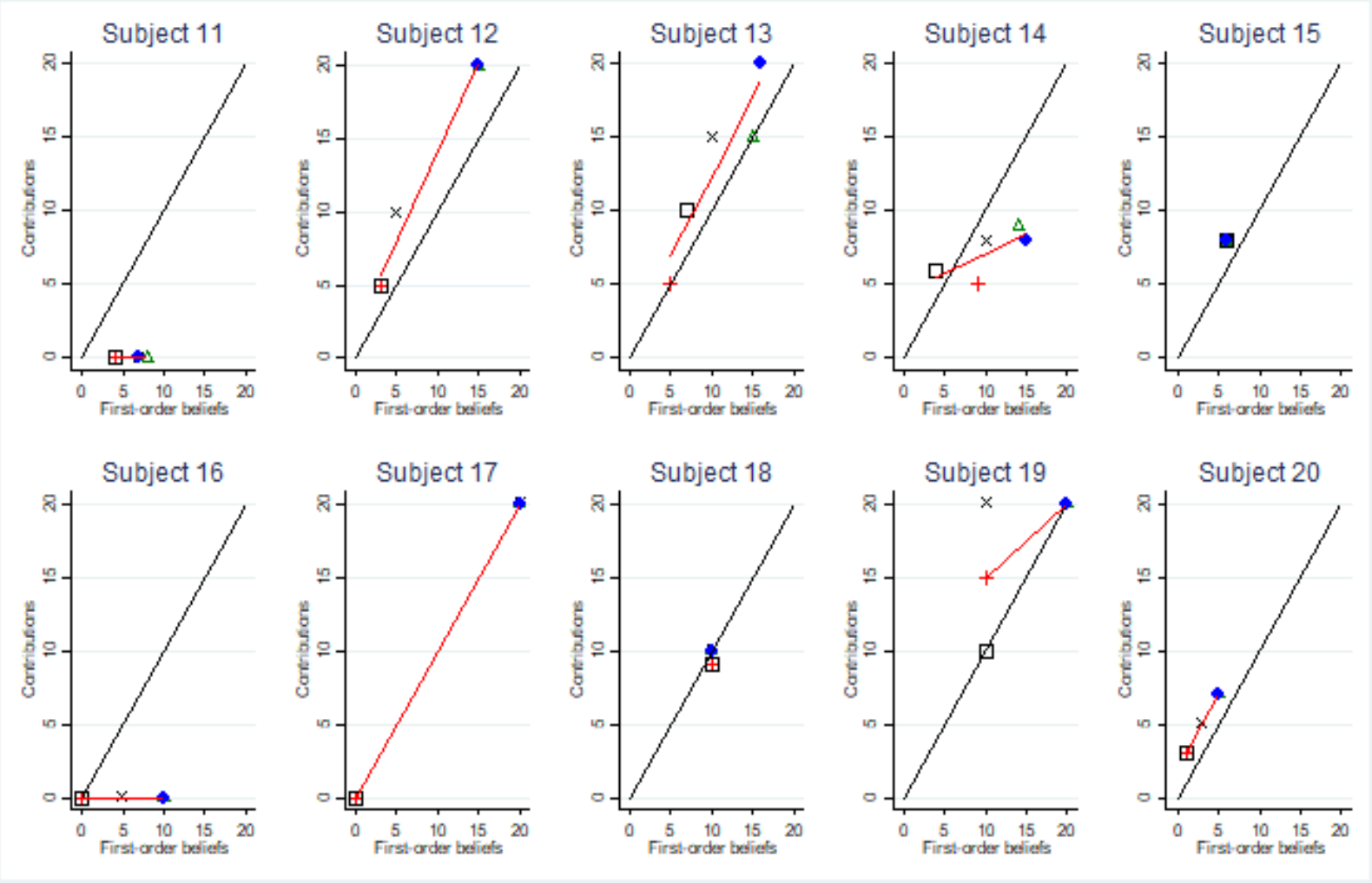
3. Individual subjects' scatter plot and assignment of types

Figure A.4 Plot of beliefs and contribution for individual subjects, with type assignments for each subject (by ID of data record) for 5-type (above, each cell) and 8-type (below, each cell) classification schemes. The best fitted lines and the 45-degree lines in red and black, respectively

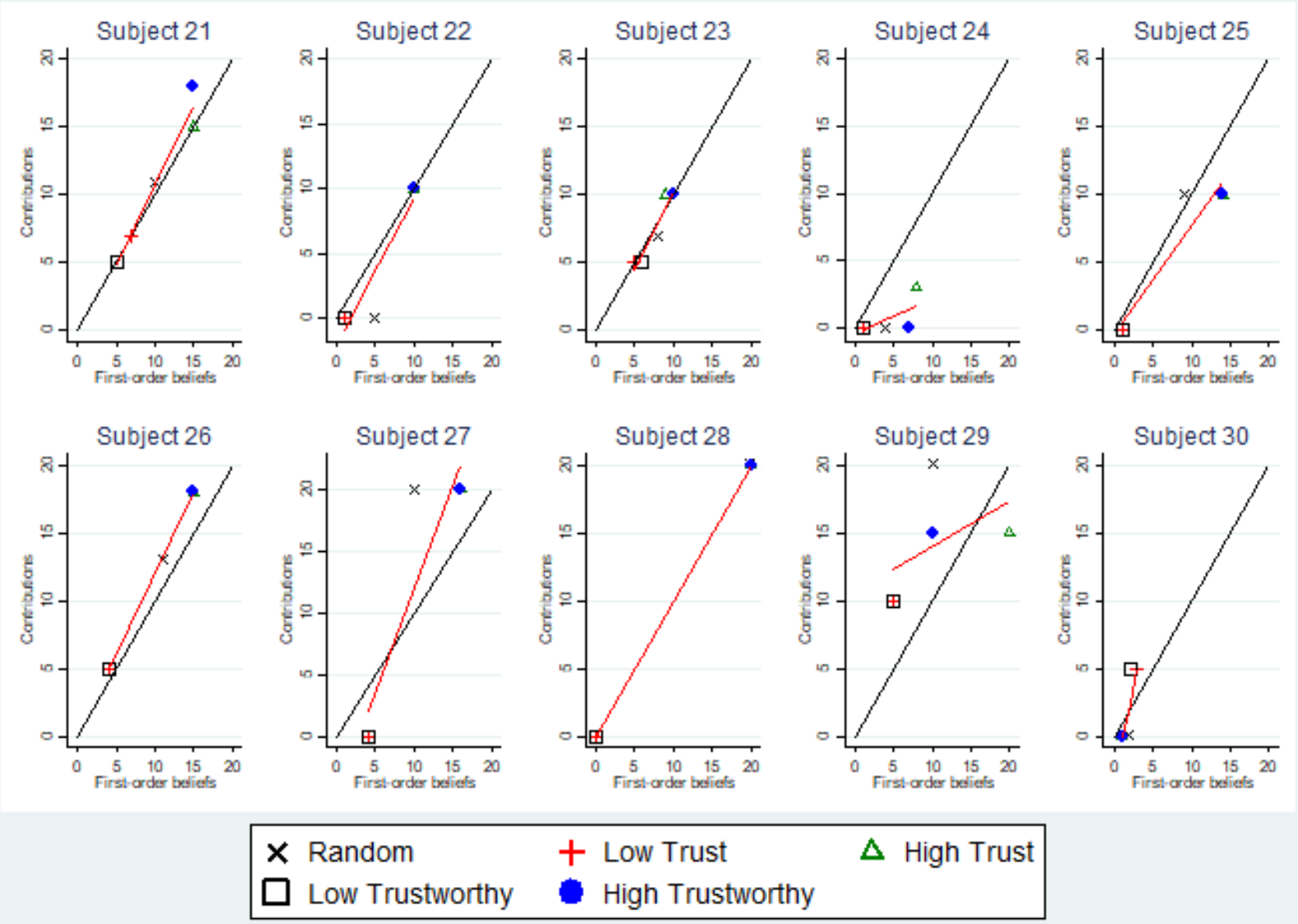
falling in a broader version of "altruist", but we did not create the category because there were no such cases among our subjects.



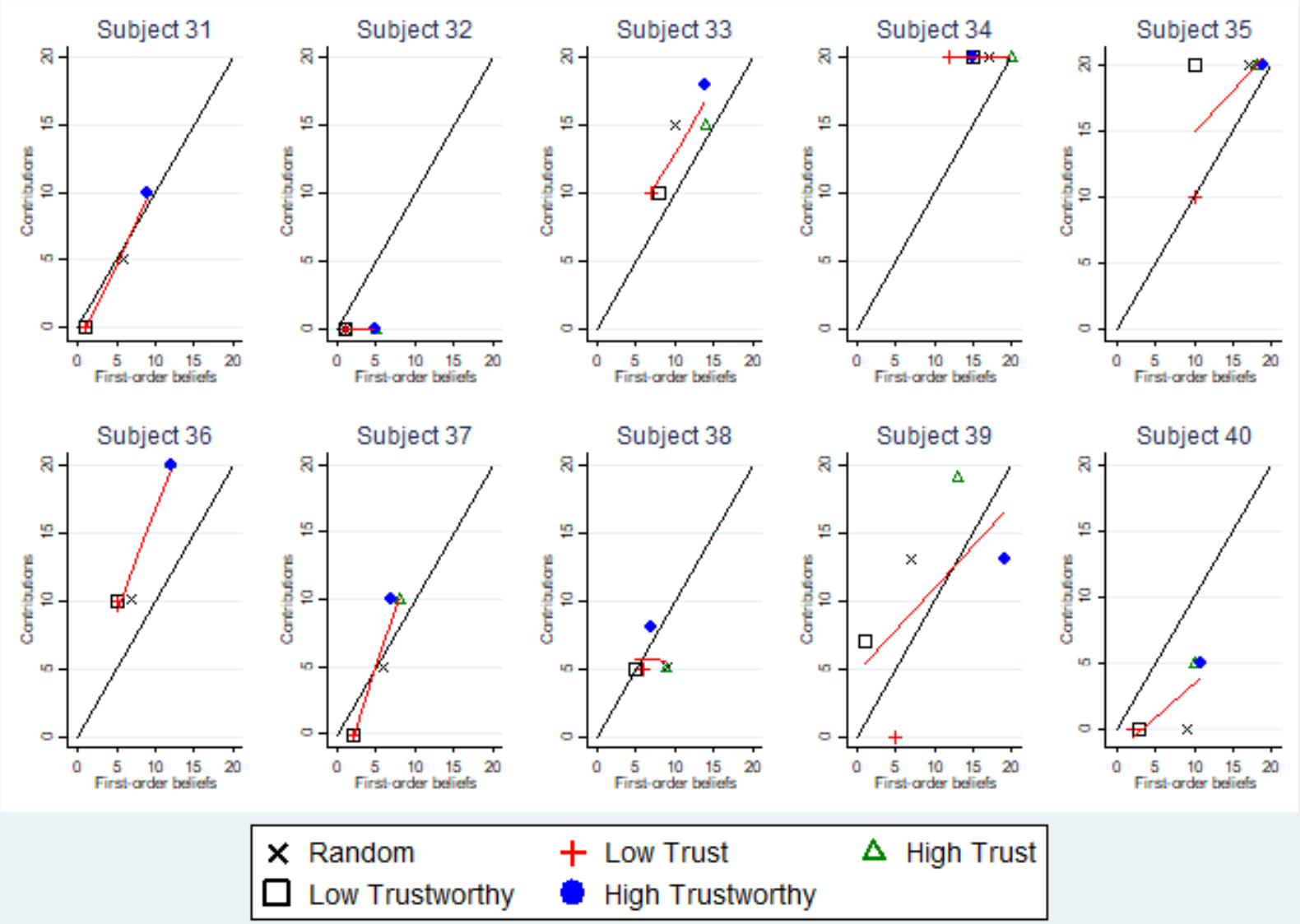
Subject	1	2	3	4	5
5 types 8 types	Other Other	Strong CC Strong CC	Strong CC Strong CC	Weak CC Weak CC	Weak CC Weak CC
	6	7	8	9	10
5 types 8 types	Weak CC Weak CC	Free rider Free rider	Weak CC Weak CC	Strong CC Strong CC	Weak CC Weak CC



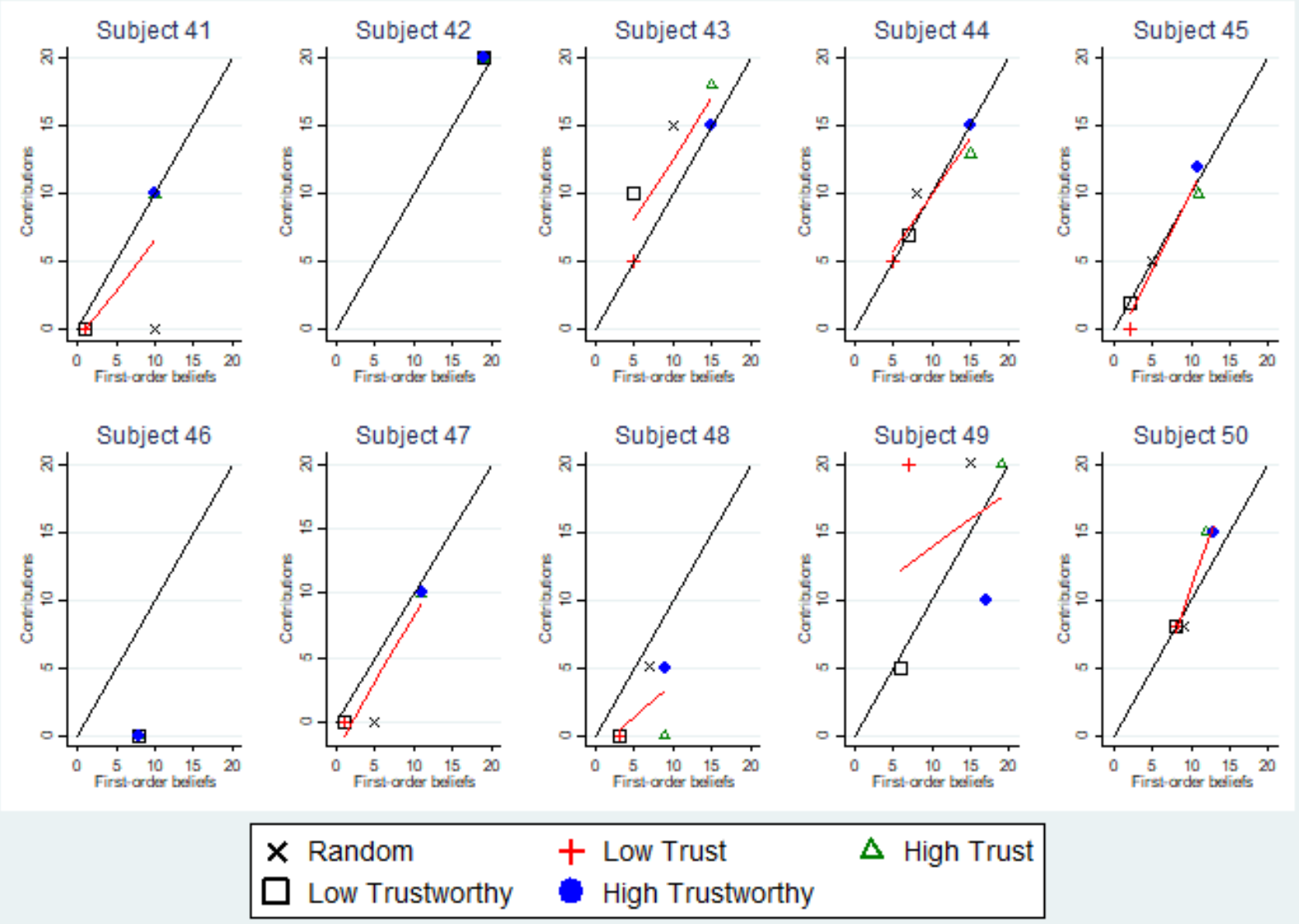
11	12	13	14	15
Free rider Free rider	Strong CC Strong CC	Strong CC Strong CC	Weak CC Weak CC	 Other Broad CC
16	17	18	19	20
Free rider Free rider	Strong CC Strong CC	Other Broad CC	 Strong CC Strong CC	Strong CC Strong CC



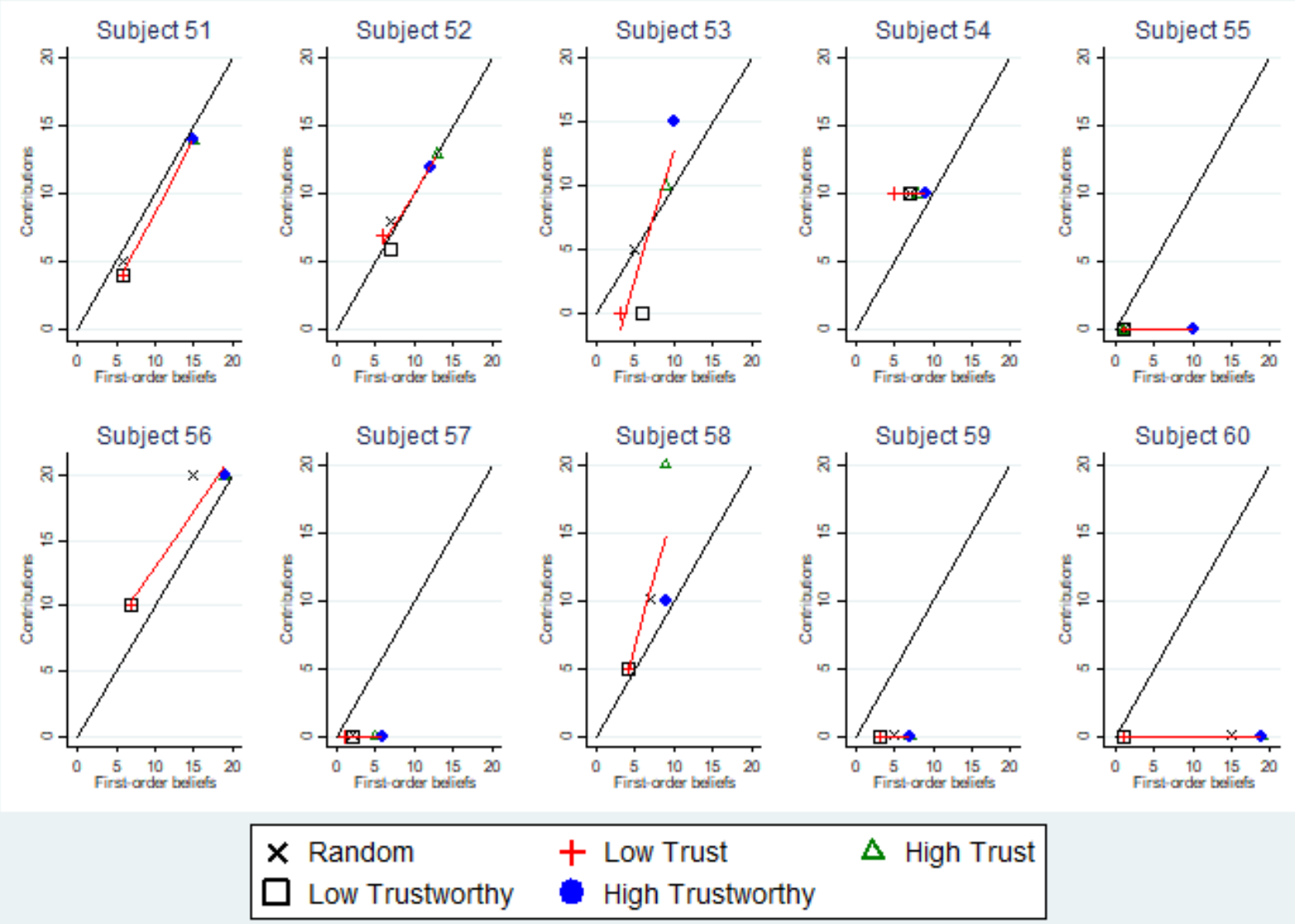
21	22	23	24	25
Strong CC Strong CC	Weak CC Weak CC	Weak CC Weak CC	Other Broad free rider	Weak CC Weak CC
26	27	28	29	30
Strong CC Strong CC	Weak CC Weak CC	Strong CC Strong CC	Strong CC Strong CC	Weak CC Weak CC



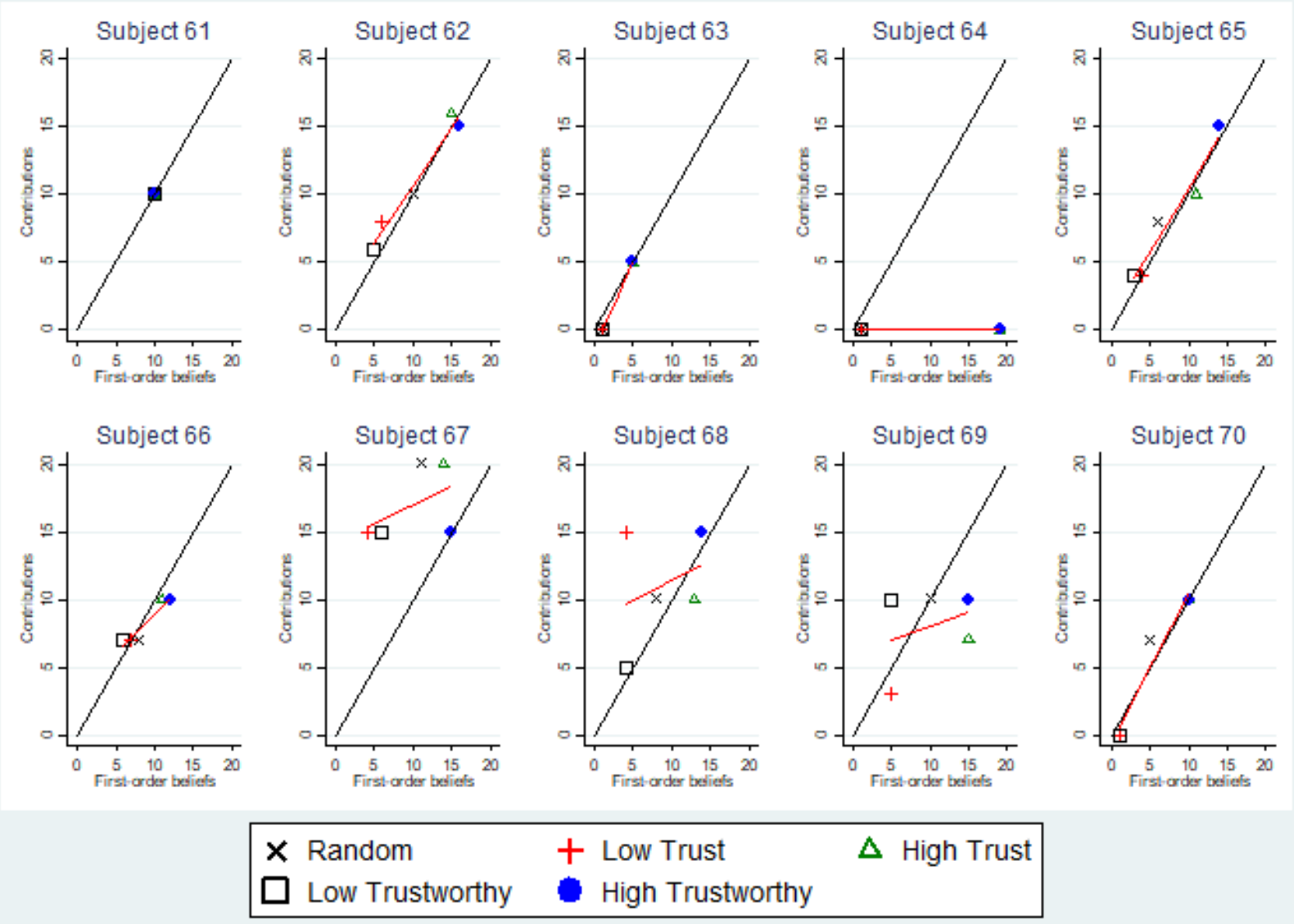
31	32	33	34	35
Weak CC Weak CC	Free rider Free rider	Strong CC Strong CC	Altruist Altruist	Strong CC Strong CC
36	37	38	39	40
Strong CC Strong CC	Weak CC Weak CC	Other Hump-shaped	Weak CC Hump-shaped	Weak CC Weak CC



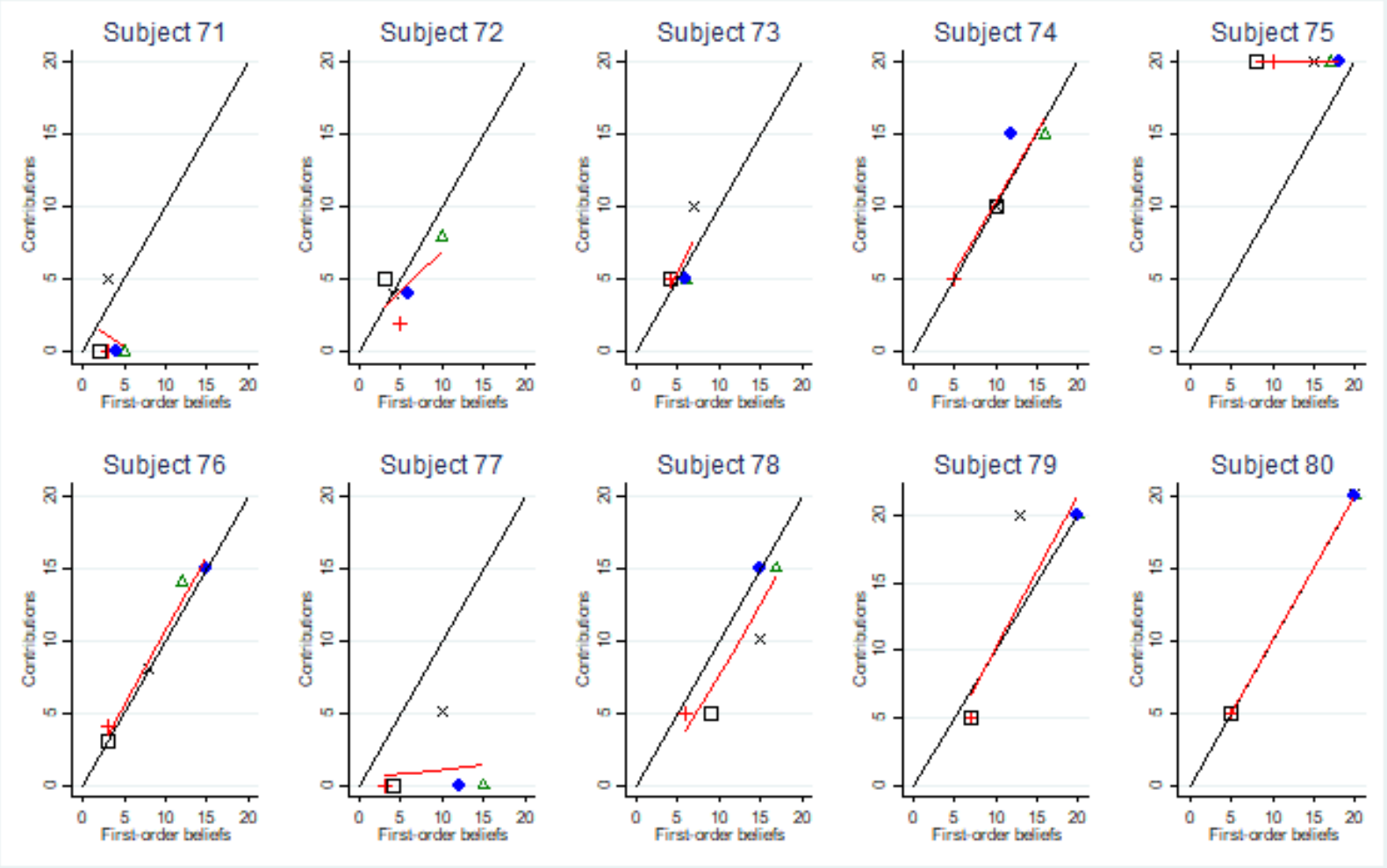
41	42	43	44	45
Weak CC Weak CC	Altruist Altruist	Strong CC Strong CC	Strong CC Strong CC	Weak CC Weak CC
46	47	48	49	50
Free rider Free rider	Weak CC Weak CC	Weak CC Weak CC	Weak CC Weak CC	Strong CC Strong CC



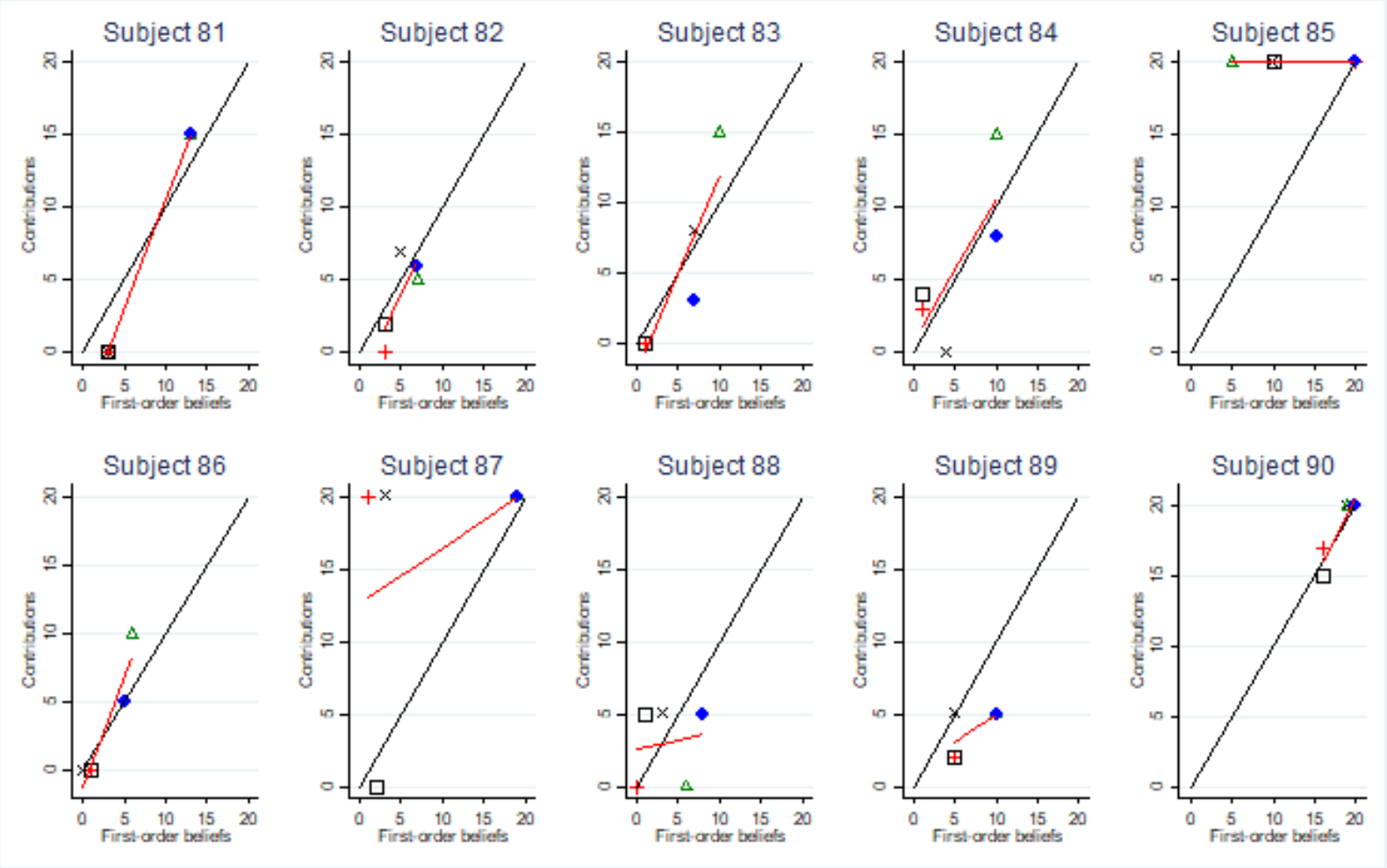
51	52	53	54	55
Weak CC Weak CC	Strong CC Strong CC	Weak CC Weak CC	Other Broad CC	Free rider Free rider
56	57	58	59	60
Strong CC Strong CC	Free rider Free rider	Strong CC Strong CC	Free rider Free rider	Free rider Free rider



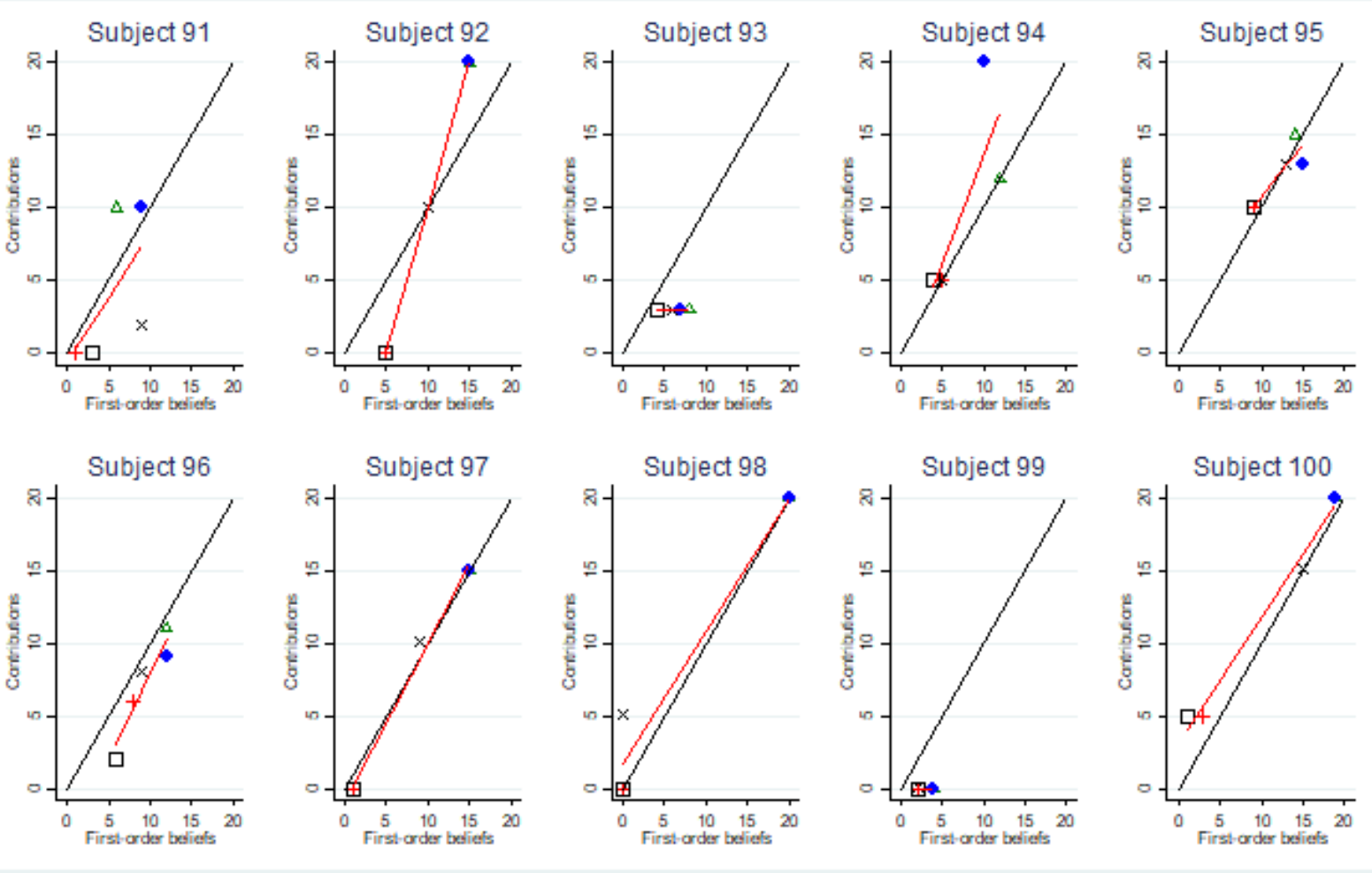
61	62	63	64	65
Other Broad CC	Strong CC Strong CC	Weak CC Weak CC	Free rider Free rider	Strong CC Strong CC
66	67	68	69	70
Weak CC Weak CC	Strong CC Strong CC	Strong CC Strong CC	Weak CC Hump-shaped	Weak CC Weak CC



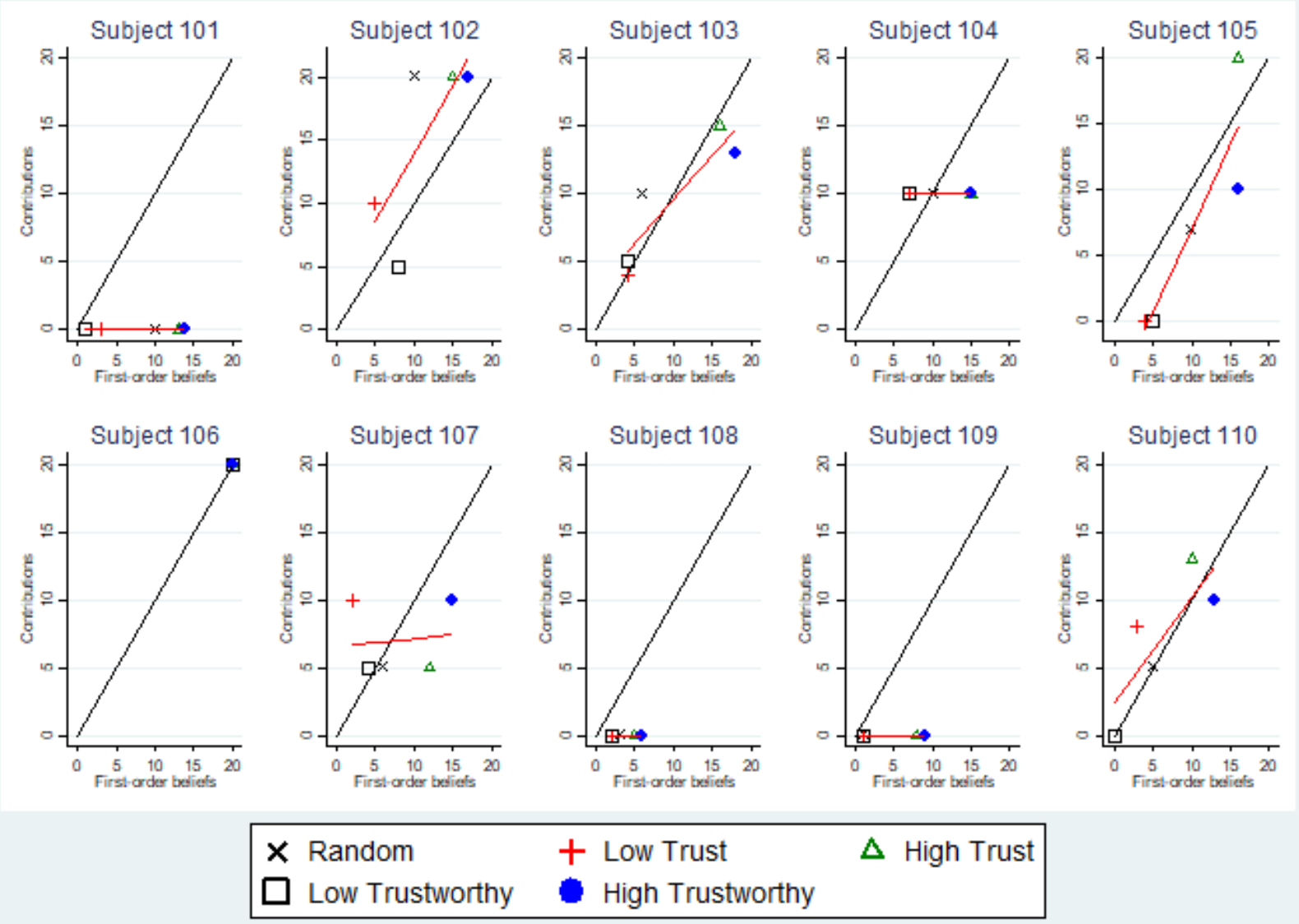
71	72	73	74	75
Other Broad free rider	Weak CC Weak CC	Weak CC Weak CC	Strong CC Strong CC	Altruist Altruist
76	77	78	79	80
Strong CC Strong CC	Other Broad free rider	Weak CC Weak CC	Weak CC Weak CC	Strong CC Strong CC



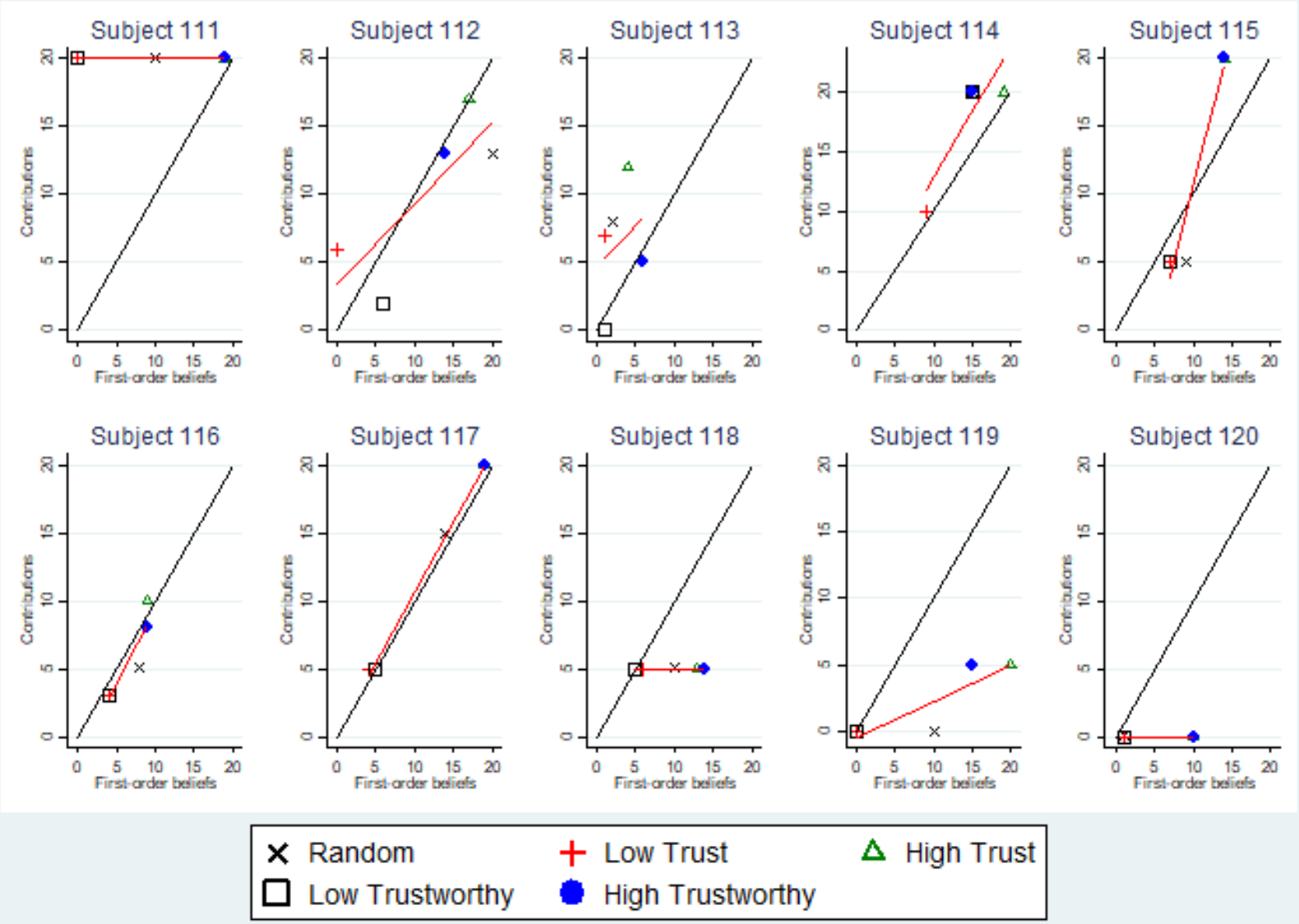
81	82	83	84	85
Weak CC Weak CC	Weak CC Weak CC	Weak CC Weak CC	Weak CC Weak CC	Altruist Altruist
86	87	88	89	90
Weak CC Weak CC	Strong CC Strong CC	Weak CC Weak CC	Weak CC Weak CC	Strong CC Strong CC



91	92	93	94	95
Weak CC Weak CC	Weak CC Weak CC	Other Other	Other Broad CC	Strong CC Strong CC
96	97	98	99	100
Weak CC Weak CC	Weak CC Weak CC	Strong CC Strong CC	Free rider Free rider	Strong CC Strong CC



101	102	103	104	105
Free rider Free rider	Strong CC Strong CC	Weak CC Weak CC	Other Broad CC	Weak CC Weak CC
106	107	108	109	110
Altruist Altruist	Weak CC Weak CC	Free rider Free rider	Free rider Free rider	Strong CC Strong CC



111	112	113	114	115
Altruist Altruist	Weak CC Weak CC	Weak CC Weak CC	Strong CC Strong CC	Weak CC Weak CC
116	117	118	119	120
Weak CC Weak CC	Strong CC Strong CC	Other Broad CC	Weak CC Weak CC	Free rider Free rider