Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Trustworthy by Convention

Maria Bigoni
Stefania Bortolotti
Marco Casari
Diego Gambetta

Quaderni - Working Paper DSE N° 827

Trustworthy by Convention

Maria Bigoni Stefania Bortolotti Marco Casari*

University of Bologna

Diego Gambetta†

Oxford University

May 10, 2012

Abstract

Social life offers innumerable instances in which trust relations involve multiple agents. In an experiment, we study a new setting called Collective Trust Game where there are multiple trustees, who may have an incentive to coordinate their actions. Trustworthiness has also a strategic motivation, and the trusters' decision depends upon their beliefs about the predominant convention with regard to trustworthiness. In this respect, the Collective Trust Games offers a richer pattern of behavior than dyadic games. We report that the levels of trustworthiness are almost thirty percentage points higher when strategic motivations are present rather than not. Higher levels of trustworthiness also led to higher levels of trust. Moreover, strategic motives appear as a major drive for trustees, comparable in size to positive reciprocity, and more important than concerns for equality.

Keywords: trust game, coordination, inequality aversion, reciprocity, experiments.
JEL codes: C92, C72, D03

1 Introduction

The standard conceptualization of trust rests on a dyadic relation, between a truster and a trustee. The archetype, which often social scientists have in mind, is an elementary

*Bigoni, Bortolotti, Casari (corresponding author): Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy; maria.bigoni@unibo.it, Phone: +390512098122, Fax +39051 0544 522, stefania.bortolotti@unibo.it, Phone: +39051 209 8135, marco.casari@unibo.it, Fax: +39051 209 8493.

†Gambetta: Nuffield College, Oxford University, New Road, Oxford, OX1 1NF, United Kingdom, diego.gambetta@nuffield.ox.ac.uk, Fax: +0044 1865278621.

market transaction between, say, a lender and a borrower: if the lender lends, the borrower must decide whether to return the money or pocket it; the lender has to guess in advance what decision the borrower will make before deciding whether to refuse or grant the loan – her most pressing question is: will the borrower be driven by his most basic self-interest or will he be able to resist that temptation and behave trustworthily? Economic life is replete with instances of this type, especially in non-simultaneous exchanges in which whoever acts first makes oneself vulnerable to being exploited: 'will the seller take my money and never send me the good? Will the buyer get the good and never send me the money?'

Social life, however, offers innumerable instances in which trust relations involve multiple agents. As the handful of authors who have studied non-dyadic cases show, matters can change quite significantly when there are multiple trusters or multiple trustees, or both (McEvily et al., 2006; Mittone and Ploner, 2009; Cassar and Rigdon, 2011; Regner and Riener, 2011).

Here we study the important and hitherto disregarded case, in which there is more than one trustee and trustees' payoffs depend on each other's actions. This case replicates the core of many real-life instances. Microcredit banks lend to groups of borrowers making them responsible for each other's solvency. In team-based research and dissemination every co-author has to trust every other to do her fair share, and acknowledge the fair share done by the others. In a mafia family, the boss entrust a "crew" of his men to mete out a violent punishment. In sport teams or military combat units, taking the lead relies on one's mates to follow suit. It is hard to score or perform heroic feats without one's comrades' assistance, and comrades need to coordinate.

It was indeed reflecting on a dismal failure of esprit de corps that first led us to embark in this research. Naples, 1982, a trial of members of a Camorra gang: in the video of the proceedings one can see boss, Pasquale D'Amico, spring up from his chair to shout and insult another defendant (Gianni Melluso) who had turned state witness. One among about ten Carabinieri, who are standing behind D'Amico's chair, acts decisively, grabs the camorrista by his shoulders and forces him to sit down. D'Amico jumps to his feet again and turns around – chest jutting, fists menacing, threats flying – to confront the Carabiniere responsible for that gesture. Instead of closing ranks and standing by their colleague who is being intimidated, most other Carabinieri ignore him, two of them try to appease the thug and two others demand an explanation, not from D'Amico but from their own colleague, as in "how could you possibly dare to rough this gentleman up!?" Unconfirmed rumors indicate that the bold Carabiniere was later kneecapped. Regardless, he was not likely to trust his colleagues again any time soon (Marco Jacquemet, personal communication).[1]

[1]Marco Jacquemet mentions briefly this episode in his "Credibility in Court: Communicative Practices in the Camorra's Trials", 1996.

The new trust game with two trustees, which we present in this paper, aims to capture the essence of situations such as those we have just described, in which the outcome cannot be reduced to the sum of the outcomes of many dyadic ones. We show both how, under conditions of payoffs interdependence, the trustees' decision of whether to be trustworthy gains an additional motivation – conforming to a convention – which cannot exist in the dyadic game; and how trusters take the additional motivation into account in deciding whether to trust.

There is an asymmetry at the core of a trust situation. While the truster's choice, whether to trust or not to trust can be driven by self-interest – 'if he pays the interest on my loan I am better off than not giving him a loan' – the trustee cannot be at the same time trustworthy and pursue his 'raw' self-interest. If the payoffs are as in the original investment game (Berg et al., 1995) and in its many followers, and if the only motivation the trustee has is to maximise monetary returns, then there is only one equilibrium, not to return the money if lent and thus, from the truster's point of view, not to lend it in the first place. To subvert this equilibrium and to trust, a truster must believe that the trustee possesses some trust-warranting property that stops him from yielding to the raw payoffs and, all things considered, makes him behave trustworthily. The primary dilemma for a truster is to which set of payoffs will the truster respond – will he follow his raw or his "all-in" payoffs and be trustworthy (Bacharach and Gambetta, 2001)?

Some all-in payoffs that induce trustworthiness can express the trustee's long-term self-interest, such as that which is satisfied by building a reputation for those who want to stay in business for an indefinite period (Dasgupta, 1988). But the most interesting trust-warranting properties typically induce a deviation from self-interest however defined – evidence of the existence of which being that even in non-repeated interactions, in which reputation cannot play a part, people show that they care about others' welfare besides their own. There are several trust-warranting properties of this kind – e.g. positive reciprocity; other-regarding preferences (Cox, 2004; McCabe et al., 2003); and trust responsiveness (Bacharach et al., 2007).[2] Some may have evolved into individual preferences and be triggered spontaneously in certain conditions – by kin or co-ethnics or needy people for instance – others may result from internalized social norms which reward kindness, fairness, or simply the respect of promises, while punishing selfishness.

When trust involves more than one trustee a new trust-warranting force can come into play, and induce even self-interested trustees to behave trustworthily. We designed a trust game in which an upright trustee motivated by unconditional trust-warranting properties will choose to reciprocate trust no matter what he believes the other trustee

[2]As reported in (Bacharach et al., 2007), trust responsiveness is a "tendency to fulfill trust because you believe that it has been placed on you" (350).

will or will not do. These are the staunch 'Kantian' trustees. But many humans act conditionally or for their self-interest. Now, the novel feature of our design is to introduce a coordination game among trustees to test whether even those who have no primary motive to act trustworthily still manage to do so. This, we hypothesize, they do if: *(i)* their final raw payoffs depend on what the other trustee will do and *(ii)* they expect that the other trustee is more likely to choose to be trustworthy than not, or if they believe that the other trustee will believe that they will be trustworthy and so on (i.e., strategic trustworthiness).

In other words, if they believe that the ensemble of trust-warranting properties present in their world is such as to incline a sufficient number of other trustees to be trustworthy, even self-interested rational trustees who would not be trustworthy in a dyadic case end up being so because their selfish payoffs are best satisfied by coordinating with what other trustees do. This is interesting for the primary motives for trustworthiness of others also impact on those who are not sensitive to them, because it is to their advantage to treat them as a convention upon which to coordinate their expectations in interactions that possess multiple equilibria (Young, 2008). Naturally the opposite can hold too: when everyone believes everyone else is more likely to be mean rather than not, anyone who is conditionally trustworthy will end up being untrustworthy. Our goal is to test whether and to what extent being '(un)trustworthy by convention' can explain subjects' behavior over and above the effects of the trust-warranting properties that make trustees deviate from their self-interest.

We report three main findings. First, trustworthiness is a stunning 30 percentage points higher in treatments where strategic motives are in place (i.e., trustees have to coordinate among themselves). Second, trust as well is sizably larger when penalties for mis-coordinations among trustees are present and trust is more profitable in these treatments. Third, by teasing apart the other trust-warranting properties, we find that strategic motives emerge as a major drive of trustworthiness, comparable in size to positive reciprocity, and larger than concerns for equality.

The structure of the paper is as follows. Section 2 presents the Collective Trust Game and relates our new game to the existing literature. Section 3 describes the experimental design and procedures. Section 4 discusses the main results of the experiment and Section 5 concludes.

2 Collective Trust

We say that a person trusts another if she acts on the expectation that the trustee(s) will do X, which is a particular task,[3] when both truster and trustee know that two conditions obtain: *(i)* if the trustee fails to do X, the truster would have done better to act otherwise; but if the trustee does X then the truster is better off than if she had acted otherwise; in other words, well-placed trust pays off, which implies that trusters who trust can be motivated by self-interest; *(ii)* the truster acting in the way she does gives the trustee the opportunity to pursue a selfish reason not to do X – in other words, the truster voluntarily puts herself in a vulnerable position, for instance by placing resources at the disposal of the trustee; 'if I hadn't lent him the money he could not have cheated me'.[4] In many but not all situations a third condition applies, namely also the trustee is *(iii)* better off being trusted than not being trusted, regardless of whether he is or is not trustworthy.[5] We say that a trustee is trustworthy if he simply does X when the above conditions apply.

The essence of trust has been captured in laboratory experiments by means of the so called investment game, also known as trust game (Berg et al., 1995). In such game, a principal (truster, hereafter) receives an endowment E and has to decide how much of this endowment to invest. The amount invested, k, is exogenously multiplied – usually by a factor $\alpha = 3$ – and transferred to an agent (trustee, hereafter). The trustee has to decide the amount, r, to return to the truster. When considering the raw payoffs of the games and not the all-in payoffs, a self-interested profit-maximizer trustee should not return any money and, by backward induction, the truster should not invest. In contrast with theoretical predictions, trust and trustworthiness have commonly been observed in the lab and this result is robust to a number of structural variations (for a review of the literature, see Johnson and Mislin, 2011; Camerer, 2003).

The level of money invested is commonly intended as a measure of the level of trust. However, the level of money invested can be explained by at least three components: *(i)* (unconditional) other-regarding preferences; *(ii)* risk attitudes; and trust proper, namely the *(iii)* beliefs about opponent's trustworthiness (see, for instance, Cox, 2004; Eckel and Wilson, 2004; Sapienza et al., 2007; Schechter, 2007; Houser et al., 2010). Analogously, the choice of returning some amount of money to the truster can be prompted by a set of

[3]The notion of trust with which we work combines elements proposed by Coleman (1990), Fehr (2009)'s concise interpretation of Coleman's, and Bacharach and Gambetta (2001). We think of trust as task-specific, largely because the motivations of the trustee to be trustworthy can be task specific: what makes one not jump a red light, steal from a neighbor who gave him the keys to feed the cat, or return a bicycle lent to him may differ, and one may trust someone in one case and not another.

[4]See also Rousseau et al. (1998): "...intention to accept vulnerability based upon positive expectations of the intentions or behavior of others"(395).

[5] This is not true of trust situations in which the trustee is merely saddled with a burden – as when he is asked, say, to look after his neighbor's cat, a case in which only *(i)* and *(ii)* obtain.

motives: *(i)* positive reciprocity, *(ii)* (unconditional) other-regarding preferences, or (iii) trust responsiveness. Cox (2004) and McCabe et al. (2003) show that positive reciprocity accounts for a large part of the observed trustworthiness, even though (unconditional) other-regarding preferences are non negligible.

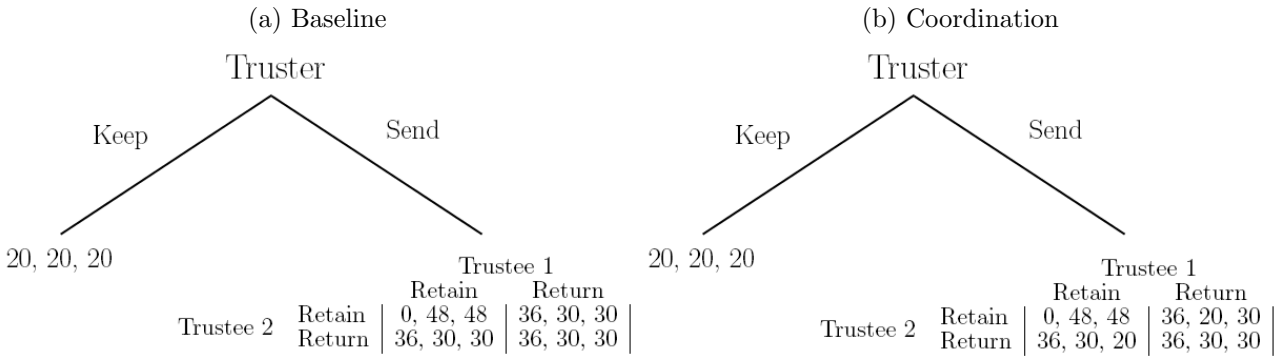
A number of scholars have extended the classic framework, by moving from dyadic to multi-players trust settings (i.e., multiple trusters or trustees). While fitting the most cited definitions of trust (for instance, see Coleman, 1990; Fehr, 2009; Bacharach and Gambetta, 2001), multi-players settings also bring about new trust-warranting properties overlooked when one employs the standard dyadic framework. Cassar and Rigdon (2011) provide evidence that trust and trustworthiness are inherently comparative and what matters the most for trustworthiness is the relative, rather than the nominal, amount received. Mittone and Ploner (2009) study the role of social effects, such as peer pressure and social spillovers, on trustworthiness in a modified investment game with multiple trustees, and find that peer pressure increases reciprocity. Social effects are also studied by Regner and Riener (2011) who provide evidence that when trustees move sequentially, the second imitates the first, but only if that is in her best interest. Finally, McEvily et al. (2006) show that trusters generalize previous experience with one member of a group to the entire group, even when group membership is created via minimal group paradigm.

Extending this line of research, our work investigates the role of payoff-interdepence among trustees by means of novel game we dub the "Collective Trust Game". The Collective Trust Game captures a situation involving one truster and two trustees with interdependent payoffs. This introduces two elements absent in the standard dyadic setting. First, trust decisions rely on beliefs about trustees' interactions and not just on beliefs about dispositions of single individuals – the *societal* dimension of trust. Second, trustees' payoffs are interdependent and the response to a trusting decision depends on the social interaction among trustees.

We employ two variants of the Collective Trust Game, *Coordination* and *Baseline* (Figure 1). In both variants, the truster and the trustees face binary decisions. The truster has to choose whether to trust (action "Send") or not (action "Keep"). If the truster chooses Keep, the two trustees have no choice to make. If the truster chooses Send, each of the trustees chooses between "Retain" and "Return". Trust increases aggregate surplus in any case, but it pays off for the truster only if at least one trustee returns. Both variants satisfy the definition of trust but differ in the way trustees' payoffs are interdependent.

In *Coordination*, when they are trusted, the trustees play a subgame similar to the stag-hunt. In the stag-hunt game players face a penalty for mis-coordination: a trustee who Retain, while the other Return earns 20 (instead of 30). The subgame played by the

Figure 1: Collective Trust Game



trustees has three Nash equilibria: a payoff-dominant equilibrium and a payoff-dominated equilibrium in pure strategies, and an equilibrium in mixed strategy. In the payoff-dominant equilibrium, both trustees choose Retain and get a payoff equal to 48, each. In the payoff-dominated equilibrium, both trustees choose Return and get a payoff of 30. In the mixed strategy equilibrium, each trustee plays Retain with probability 9/14.

For the truster trust always pays off, unless trustees coordinate on the payoff-dominant equilibrium, (Retain, Retain). The break-even point in expected terms is when each trustee returns with an independent probability of 1/3. This value yields for the truster a payoffs of 0 with probability 4/9 and a payoff of 36 with probability 5/9.

In other words, the *Coordination* variant of the Collective Trust Game admits multiple subgame perfect Nash equilibria, and the outcome depends on the social convention in place as equilibrium selection criterion. If the predominant social convention among trustees prescribes taking advantage of people whenever possible then the truster is worse off trusting than not trusting. If instead the social convention inclines the trustees either towards the mixed strategy equilibrium, or prescribes being trustworthy, then the truster is better off trusting. This game bears some resemblance to the Collective Resistance Game employed in Political Science (Weingast, 1997; Cason and Mui, 2007), which differs from the Collective Trust Game in that the first mover ("leader") has the chance to increase her own payoff by reducing (rather than increasing) the social welfare.

Baseline differs from the *Coordination*, because in the former there is no penalty for mis-coordination of actions among trustees (Figure 1). Baseline is the closest to the standard trust game with two agents. The subgame between trustees has two Pareto-ranked Nash equilibria in pure strategies. As in *Coordination*, in one equilibrium both trustees choose Retain and individual payoffs are 48, while in the other equilibrium both trustees choose Return and individual payoffs are 30.

The equilibrium (Send,Return,Return) is "weak" as it does not withstand a refinement based on evolutionary arguments (see proof in the Appendix). Even though *Baseline* nom-

inally admits multiple subgame perfect Nash equilibria, there exists a unique evolutionary stable equilibrium compatible with our subjects' beliefs (as elicited before period 1 of the stage game). As before, trust pays off in expected terms if each trustee chooses Return with an independent probability of 1/3. Hence, trust does not pay if the trustees coordinate on the evolutionary stable equilibrium.

This design allows us to capture what we labeled the strategic components of trust and trustworthiness. We say that a trustee's choice to Return is driven by "strategic trustworthiness" if it is self-regarding and depends only on the belief that other trustees will also choose Return. Notice that this definition does not include herd behavior while it well suits the definitions of trust proposed in the literature. We say that a truster's choice of Send is driven by "strategic trust" if it is motivated by the belief that trustees will be strategically trustworthy. In *Baseline* (and of course in the dyadic trust game), neither strategic trust nor trustworthiness play a role. A comparison of *Baseline* and *Coordination* allows us to isolate the strategic component in trust and trustworthiness from other determinants and assess the relative importance of each.

3 Experimental Design

We implemented a two by two between-subjects factorial design with four treatments (Table 1). One dimension of variation concerns the way trustees' payoffs are interdependent in the Collective Trust Game. In the *Coordination* treatments there is a penalty for mis-coordination, while no such a penalty is present in the *Baseline* treatments (Figure 1). The other dimension concerns whether the choices Send or Keep are intentionally made by participants in the role of trusters (*Active* treatments) or they depend on a random draw (*Passive* treatments). In both cases, the decision is payoff relevant for the truster. In the *Passive* treatments, the software was programmed to draw the outcome of the first node – Send or Keep – from a period-specific probability distribution replicating the empirical distribution of choices made in the same period of all sessions of the corresponding *Active* treatment.

The four treatments elicit a different mix of trust-warranting motives. In the *Baseline-Active* treatment, trustees can choose Return for two main reasons: positive reciprocity or aversion to inequality. Hence, either they want to reciprocate a kind and intentional action of the truster, or they care about equality and want to reduce the differences in earnings between themselves and the truster. In the *Coordination-Active* treatment an additional motive is present: strategic trustworthiness. This motivation stems from the need to avoid the loss carried by mis-coordination; that is, self-interested trustees may choose Return simply because their selfish payoffs are best satisfied by coordinating. A comparison between these treatments allows us to address our two main questions: Does trustworthi-

Table 1: Experimental treatments

	Baseline Active	*Coordination Active*	*Baseline Passive*	*Coordination Passive*
Stage Game	Baseline	Coordination	Baseline	Coordination
Truster	Human	Human	Robot	Robot
Motives for Trustworthiness				
Inequality aversion	✓	✓	✓	✓
Reciprocity	✓	✓		
Strategic Trustworthiness		✓		✓
Session	03/31;04/04	03/30;04/01	04/15	04/15
Subjects	60	57	30	27
Indep. Obs	4	2	4	2

Notes: Sessions conducted in 2011. In *Baseline-Passive*, because of a technical problem, the average frequency of trust was 53 percent, which is slightly different from the average frequency of 47 percent observed in *Baseline-Active*.

ness have a strategic component? Does trust have a strategic component? Finally, in the *Passive* treatments reciprocal motives are removed on the assumption that one cannot be grateful to a truster since the decision was made by a machine, while the other trustworthiness properties remain. In the *Baseline-Passive* treatment equality-concerns are the only possible explanation for being trustworthy.[6] In the *Coordination-Passive* treatment also strategic motives come into place due to coordination concerns. A comparison between our four treatments allows us to disentangle the impact of strategic trustworthiness, from the impact of positive reciprocity and aversion to inequality.

Each session, in addition to a repeated Collective Trust Game, included choices over lotteries and dictator games. The instructions were distributed before each part and were read out aloud by the experimenter (a copy of the instructions is in the Appendix). To start with, subjects had to choose their most preferred lottery from the list of six options in Table 2. Each lottery had a low and a high payoff outcome that always occurred with a fifty percent chance (Eckel and Grossman, 2002, 2008). Only two subjects per session received a payment for the choice over lotteries. Payoffs for the lottery task were presented in Euros.[7]

In this paper we do not present the analyses of the dictator games choices and hence we refer the interested reader to the instructions in the Appendix.[8] After the lottery and

[6]Recall that in the *Passive* treatments, while trust decisions are made at random, there are in the room subjects in stand-by who earn the payoffs of the truster. Hence, the two trustees' decisions in a group impact their earnings as well as the earnings of one other person in the same way as in the *Active* treatments.

[7]A volunteer performed two manual draws from a bag with numbered balls to randomly select these subjects. An additional manual draw selected the outcome for the lottery chosen by the two subjects. To avoid carry over effects, the draws were performed at the end of the session.

[8]The three-person dictator games were employed to elicit static preferences for equality, efficiency, and self-interest. We gave no feedback information about choices and earnings until the end of the session, to minimize

Table 2: Risk aversion task

	High payoff (orange ball)	Low payoff (white ball)	CRRA coefficient	Frequency of choices
Lottery 1	17.5	17.5	>3.64	13%
Lottery 2	22.5	15.0	3.46-1.16	24%
Lottery 3	27.5	12.5	1.16-0.70	18%
Lottery 4	32.5	10.0	0.70-0.50	27%
Lottery 5	37.5	7.5	0.50-0	16%
Lottery 6	44.0	1.0	<0	2%

Notes: The Coefficient of Relative Risk Aversion (CRRA) was not shown to subjects.

dictator game choices, subjects participated in a Collective Trust Game for 30 periods. A quiz to ensure full understanding was administrated before this part; everyone had to correctly answer all questions before proceeding.

Before each period of part three, new groups of $N = 3$ were formed according to a strangers matching protocol. We partitioned participants in each session into two matching sets in a way that subjects were never matched with members outside their set; we thus have two independent observations per session. The truster and trustee roles changed at random from period to period. Changing roles facilitates learning and helps spreading norms of trust within the population; some studies pointed out that playing both roles may reduce both the levels of trust and trustworthiness (Burks et al., 2003; Johnson and Mislin, 2011). If the truster chooses Send, then the trustees had to simultaneously choose either Retain or Return. If the truster chose Keep, then the trustees had no decision to make. In the *Passive* treatments, in each period the trustees observed a draw from an urn containing blue and yellow balls. The draw selected the first move done by the computer for the passive person who was sitting in for the truster. Trustees had to make a decision only in case a blue ball was drawn. Subjects did not receive any information about the composition of the urn – which changed from period to period – or how the composition was determined. However, we made clear to the subjects that the composition was determined in advance and was unaffected by their choices during the game.

At the end of each period, subjects could observe everyone's choice and earnings within their group. The interaction was anonymous, hence there was no scope for building an individual reputation. Before periods 1 and 30, we elicited subjects' beliefs about trust and then about trustworthiness in the population. There was no payment associated with belief decisions to avoid possible interferences with choices in the trust game (Croson, 2000).[9] Earnings for the trust game were expressed in tokens. At the end of the session,

possible spill-over effects on subsequent decisions.

[9]Subjects answered the following questions before knowing their role for the period: "Out of 10 role A people

tokens were converted into Euros, at the rate of 1 Euro per 40 tokens. At the end of the Collective Trust Game part, subjects answered to a questionnaire to elicit participants' socio-demographic characteristics.

The experiment involved 174 subjects, divided in 6 sessions (Table 1) and was conducted at the Bologna Laboratory for Experiments in Social Sciences (BLESS). Subjects were mostly students at the University of Bologna, and were recruited through ORSEE (Greiner, 2004). About 53 percent of the subjects were male; nobody took part in more than one session. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). Upon arrival, subjects were randomly assigned to a cubicle to avoid eye contact, and no communication was allowed during the experiment. The average session lasted about 1 hour and 20 minutes. Subjects were paid privately in cash at the end of the session and earned on average 23 Euros.

4 Results

This section illustrates the main results and is structured around three parts. In Section 4.1 we study the *Active* treatments and identify the strategic motives for trustworthiness and trust (Result 1 and 2). In Section 4.2 we investigate the impact of beliefs and risk attitudes on trust (Results 3 and 4). Finally, in Section 4.3, we juxtapose data from *Active* and *Passive* treatments to analyze the determinants of trustworthiness. More specifically, we measure the relative weight of strategic trustworthiness in the decision to Return as compared to preferences for equality and positive reciprocity (Result 5).

Table 3: Trust and trustworthiness by treatment

	Baseline Active	*Coordination Active*	*Baseline Passive*	*Coordination Passive*
Trust	53%	81%	–	–
Trustworthiness	36%	62%	12%	50%
(Retain,Retain)	*41%*	*17%*	*77%*	*24%*
(Retain,Return)	*47%*	*41%*	*22%*	*52%*
(Return,Return)	*12%*	*42%*	*1%*	*24%*
N. Obs.	1800	1710	900	810
N. Subjects	60	56	30	27

Notes: Average level of trust and trustworthiness, all periods pooled. Trust is profitable for the truster if at least one of the two trustees decides to play Return.

[trusters], on average, how many role A people do you expect to Keep and how many to Send?" and "Out of 10 role B people [trustees] that have to take a decision, on average, how many role B people do you expect to Retain and how many to Return?"

4.1 Strategic trust and trustworthiness

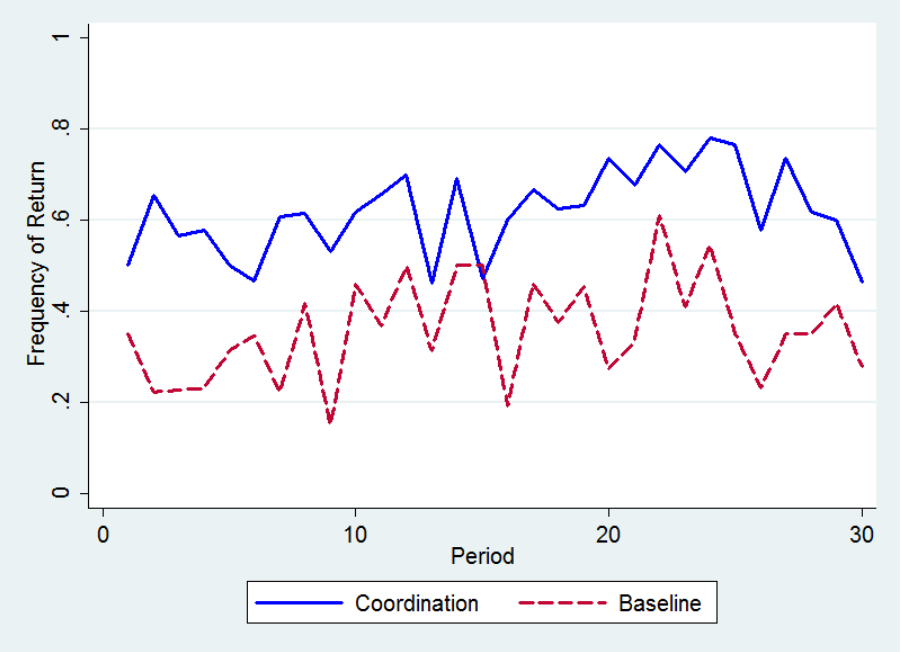
The main difference between *Baseline* and *Coordination* lies in the way trustees' choices are interdependent (summary statistics for *Baseline* and *Coordination* are reported in Table 3). Since the present manipulation has a direct effect on trustworthiness and only indirectly affects trust (i.e., trusters expect – or experience – different levels of trustworthiness), we first analyze trustees' behavior and then move to trusters' behavior.

Does trustworthiness have a strategic component? In *Coordination-Active*, trustworthiness can be induced by strategic motives, which stem from trustees' desire to avoid the cost of mis-coordination. Strategic trustworthiness can hence be measured by comparing *Coordination-Active* with *Baseline-Active* where no such motives are in place.

Result 1 *Trustworthiness has a large and significant strategic component.*

As shown in Figure 2, strategic trustworthiness plays an important role already in the first period: 50 percent of the trustees choose Return in *Coordination-Active* while only 35 percent do so in *Baseline-Active* (p=0.350, chi-squared test, N=20 and N=18). As the experiment progresses, the gap between the two treatments increases, and the relative importance of strategic trustworthiness becomes larger. Overall, average (median) trustworthiness was 0.36 (0.15) and 0.62 (0.67) in *Baseline-Active* and *Coordination-Active* respectively, and the difference is significant, albeit weakly (p=0.083, two-tailed Wilcoxon rank-sum test, N=4 and N=4).

Figure 2: Trustworthiness across periods: Active treatments



Further support for Result 1 is provided through a series of probit regressions (see

Table 4).[10] The dependent variable indicates whether the trustee chose Return (1) or Retain (0) in each period. In Model 1 the only explanatory variable is a dummy for *Coordination-Active* treatment, which is meant to capture the effect of strategic trustworthiness. According to results from this regression, the probability that a trustee chooses Return is 26 percent higher in *Coordination-Active* (Coordination in the regression) than in *Baseline-Active* (Baseline in the regression), lending support to Result 1. Variants of this model deliver qualitatively and quantitatively similar results. In Model 2, we include two additional explanatory variables which capture learning effects. In our repeated interaction setting, learning has an upper bound (i.e., the average level of trustworthiness cannot be larger than 1) and tends to be more pronounced in earlier rather than later periods. We therefore opted for the reciprocal of the variable *Period* to account for it. Interactions between 1/*Period* and the treatment dummies are included in the second model of Table 4. The negative coefficients of these interaction terms indicate that trustworthiness increases over time; however, the upward trend is statistically significant for *Coordination-Active* only. Result 1 is robust also when controlling for a number of socio-demographic characteristics collected through a questionnaire (Models 3 and 4).[11]

Table 4: Treatment effect on trustworthiness: Active treatments

Dep. var.:	Model	Model	Model	Model
Return(1) Retain (0)	(1)	(2)	(3)	(4)
Coordination (d)	0.266**	0.268**	0.252**	0.249**
	(0.113)	(0.111)	(0.117)	(0.115)
1/Period x Baseline		-0.149		-0.187
		(0.156)		(0.149)
1/Period x Coordination		-0.176***		-0.173***
		(0.033)		(0.033)
Demographics	No	No	Yes	Yes
N.obs.	1554	1554	1554	1554
Pseudo-R2	0.050	0.052	0.077	0.079

Notes: Marginal effects from a probit regression: standard errors robust for clustering at the group level (in parentheses).

We now turn to truster beahvior: *Does trust have a strategic component?*

[10]Regression models for binary data are adopted throughout the paper since truster and trustees had to make a binary choice (i.e., Send/Keep and Return/Retain). Unless otherwise specified, the paper reports results from probit regressions with standard errors robust for clustering at the matching-group level; moreover, results are robust to different specifications (i.e., logit models, probit models with individual random effect).

[11]The following socio-demographic characteristics are included in the present and subsequent regressions. *Male* is a dummy equal to 1 when the subject is a male and to 0 when female, *Age 26+* takes value 1 if the subject is 26 or older, and in the variable *Siblings* 1 indicates no siblings, and 4 indicates 3 or more siblings. Two proxies for wealth are also included: *Cars 2+* is a dummy equal to 1 if the family owns two or more cars and *House of property* is 1 if the family owns a house. Finally, we include controls for the region of origin *Center* and *South-Islands*. We considered the region where a subject attended elementary school rather than the birth place to construct the present dummies. North, Center, and South-Islands are defined following the Italian Institute of Statistics (Istat) categorization.

Result 2 *Trust has a large and significant strategic component.*

The results suggest a positive answer to our second research question, on whether strategic trust does matter: the average (median) level of trust over the 30 periods is 0.53 (0.59) in *Baseline-Active* and 0.81 (0.91) in *Coordination-Active*. Overall, trust levels are significantly different between the two conditions at 5 percent level, as revealed by a two-tailed Wilcoxon rank-sum test (p=0.021, N=4 and N=4). As illustrated in Figure 3, the initial level of trust is similar across treatments and becomes significantly different only after some experience of the game. In the initial period trust is indeed statistically indistinguishable across treatments (average trust level in period 1, is 0.50 and 0.47 in *Baseline-Active* and *Coordination-Active*, respectively; p=0.869, chi-squared test, N=20 and N=19). However, after period 1 the gap between the two treatments progressively widens; specifically, a pronounced upward trend emerges when strategic motives are present.

Figure 3: Trust across periods: Active treatments

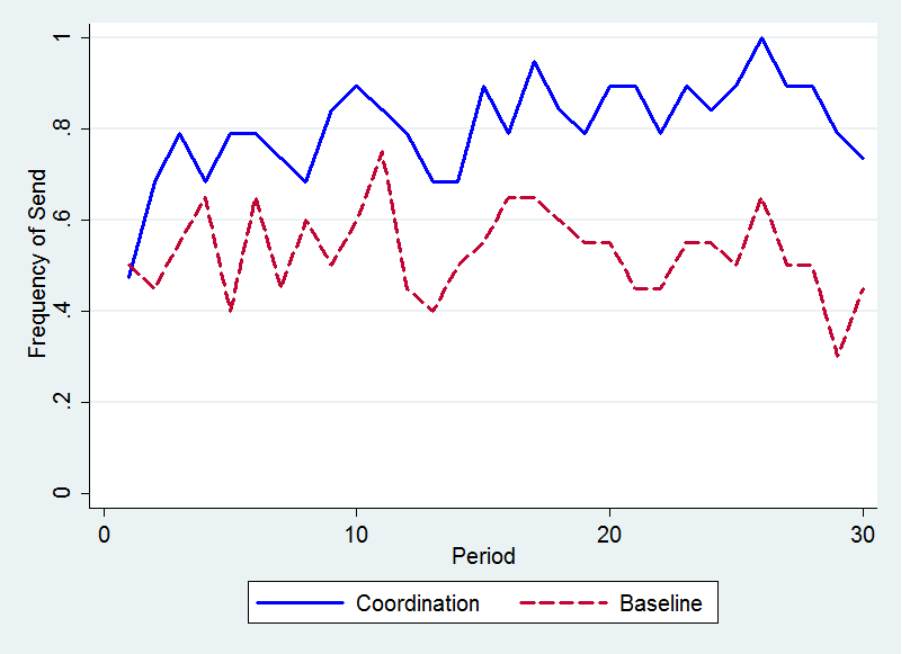


Table 5 reports results from four probit regressions providing further support to Result 2. The dependent variable is whether the truster chose Send (1) or Keep (0) in each period and the explanatory variables are the same as in Table 4. The estimated probability of observing a trustful decision in *Coordination-Active* is between 27 and 32 percent larger than in *Baseline-Active*, all else held constant. Overall, trusters chose Send more often in *Coordination-Active* than in *Baseline-Active*: the result is highly significant and robust to different specifications and to socio-demographics controls. A strong learning dynamic emerges in *Coordination-Active*: the probability of trusting increases over time, as indicated by the negative coefficient for $1/Period \times Coordination$

Table 5: Treatment effects on trust: Active treatments

Dep. var.:	Model	Model	Model	Model
Send(1) Keep (0)	(1)	(2)	(3)	(4)
Coordination (d)	0.275***	0.327***	0.279***	0.325***
	(0.061)	(0.077)	(0.064)	(0.082)
1/Period x Baseline		-0.025		-0.046
		(0.174)		(0.176)
1/Period x Coordination		-0.413***		-0.393***
		(0.122)		(0.125)
Demographics	No	No	Yes	Yes
N.obs.	1170	1170	1170	1170
Pseudo-R2	0.068	0.079	0.100	0.110

Notes: Marginal effects from a probit regression: standard errors robust for clustering at the matching-group level (in parentheses). Active treatments only.

4.2 Beliefs and risk attitudes

Is there a correlation between trust and pre-play beliefs about others' trustworthiness? That is, do subjects correctly anticipate the role of strategic trustworthiness? To assess whether trusters correctly anticipate that trustees return more when penalties for miscoordination are in place, we look at beliefs about others' trustworthiness as reported by our subjects before playing the Collective Trust Game. For simplicity, in the present discussion, we limit our analysis to beliefs about trustworthiness (beliefs, hereafter).

Result 3 *Subjects do not correctly anticipate the differences in trustworthiness across treatments.*

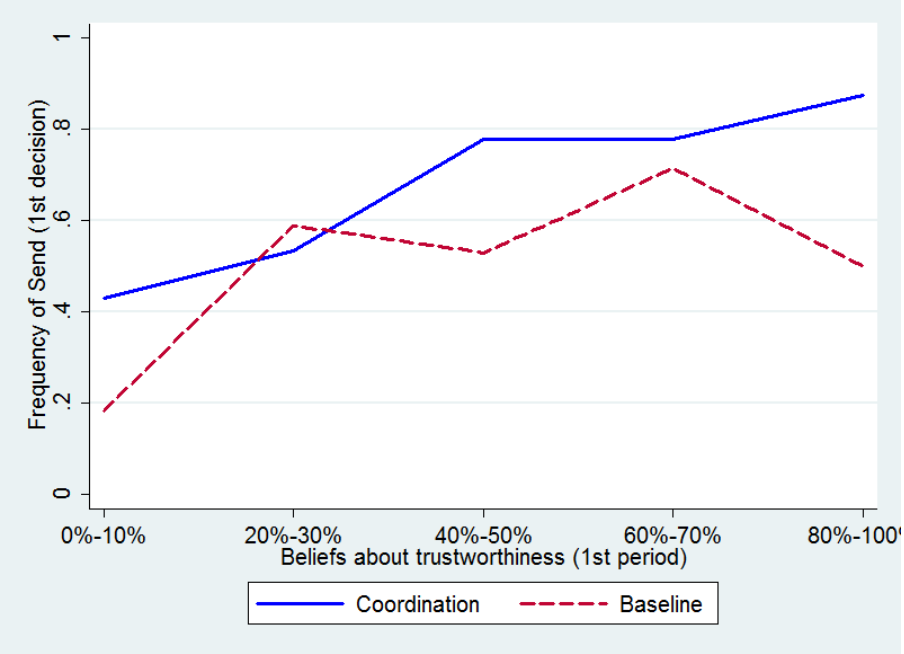
Whereas observed levels of trustworthiness are significantly different across treatments, pre-play beliefs are not: on average, subjects expected 39 percent of the trustees to play Return in the *Baseline-Active* and 44 percent in the *Coordination-Active* treatment (p=0.278, two-sample Wilcoxon rank-sum test, N=60 and N=57, two-sided). The evidence suggests that subjects did not correctly anticipate behavioral differences between the two treatments.[12]

One can also compare beliefs and actions at the individual level. As one would have expected, we find that the higher the pre-play belief about others' trustworthiness, the the

[12]Sapienza et al. (2007) maintains that beliefs are subject to updating as the game progresses due to learning; in line with this conjecture, our data show a marked difference between beliefs elicited in the first and the last period in the *Coordination-Active* treatment. Before the final period, the difference in beliefs between treatments is substantial and significant: on average subjects expected 34 percent of the trustees to play Return in the *Baseline-Active* and 58 percent in the *Coordination-Active* treatment (p=0.043, two-sample Wilcoxon rank-sum test, N=4 and N=4, two-sided). The stability of beliefs over time in the *Baseline-Active* treatment should not come as a surprise since beliefs were already accurate in the first period. It is worth noticing that beliefs in the last period and actual behavior were remarkably close in both treatments; that is, subjects correctly learned norms and conventions in use.

higher the initial level of trust (see Figure 4). To study the coherence between pre-play beliefs and initial trust behavior, we restrict our analysis to the first choice a subject made when playing as truster. In *Coordination-Active*, subjects who first chose Send expect a larger fraction of the trustees to Reciprocate as compared to subjects who first chose Keep and the difference is statistically significant according to a two-tailed Wilcoxon rank-sum test (p=0.019, N=39 and N=18); the same pattern, albeit not significant, was observed also in *Baseline-Active* (p=0.197, N=30 and N=30). Pre-play beliefs are also strongly correlated with initial trustworthy choices; while in *Coordination-Active* this positive correlation can be rational – there is a monetary cost associated to mis-coordination – the same is not true for *Baseline-Active*. In this case, the difference is higly significant for both *Coordination-Active* (two-tailed Wilcoxon rank-sum test, p¡0.001, N=32 and N=25) and *Baseline-Active* (two-tailed Wilcoxon rank-sum test, p¡0.001, N=37 and N=23).

Figure 4: Initial actions and pre-play beliefs: Active treatments



Is there a correlation between trust and risk attitudes? We found that beliefs play an important role in shaping initial behavior, but other forces, such as risk preferences, may influence trusters' choices, as well.

Result 4 *Risk attitudes significantly correlate with trust in the Baseline and not in the Coordination treatment. Pre-play beliefs significantly correlate with trust in Baseline and not in Coordination.*

To shed further light on the determinants of trust, we ran a regression on trusters' choices against their risk attitudes and beliefs. Table 6 presents probit estimations separately for *Coordination-Active* and *Baseline-Active*. Two dummies for risk attitudes and

Table 6: Beliefs and risk aversion on trust: Active treatments

Dep. var.:	Coordination		Baseline	
Send(1) Keep(0)	Model 1	Model 2	Model 3	Model 4
1/Period	-0.308***	-0.288***	-0.028	-0.060
	(0.088)	(0.070)	(0.233)	(0.249)
Strongly risk averse (d)	-0.041	-0.054	-0.101***	-0.151*
	(0.054)	(0.075)	(0.038)	(0.088)
Risk neutral or loving (d)	-0.006	-0.028	0.305***	0.250***
	(0.046)	(0.082)	(0.079)	(0.093)
Low belief (p=1) (d)	-0.060	-0.072	-0.194***	-0.128*
	(0.102)	(0.106)	(0.069)	(0.076)
High belief (p=1) (d)	0.025	0.006	0.069	0.122***
	(0.057)	(0.061)	(0.060)	(0.040)
Demographics	No	Yes	No	Yes
N.obs.	570	570	600	600
Pseudo-R2	0.038	0.077	0.079	0.109

Notes: Marginal effects from a probit regression: standard errors robust for clustering at the matching-group level (in parentheses). Active treatments only.

for beliefs are included in all specifications. The dummy *Strongly risk averse* (*Risk neutral or loving*) takes value 1 for subjects whose choices in the lottery task is compatible with a relative risk aversion coefficient smaller than 0.50 (larger than 1.16), and 0 otherwise. Similarly, we created two dummies for pre-play beliefs on others' trustworthiness; *Low belief* (*High belief*) is equal to 1 if a subject expects less than 1/3 (more than 2/3) of the trustees to be trustworthy.

In *Baseline-Active*, unlike in *Coordination-Active*, higher levels of trust are associated with more risk tolerance, and the correlation is statistically significant (Table 6). Even tough pre-play beliefs do not seem to play a significant role in the *Coordination-Active* treatment, they do so in the *Baseline-Active* treatment, where they can explain a rather large proportion of the variation of trust across subjects. These two findings suggest that risk attitudes are relevant only when returns from trustful actions are small and playing Send is highly risky (i.e., *Baseline-Active*).

4.3 Decomposing trustworthiness

What is the relative contribution of strategic trustworthiness, positive reciprocity, and concerns for equality? We carry out a comparison of Return frequencies across treatments in order to disentangle the motivations behind trustworthiness into (i) strategic concerns to avoid mis-coordination, (ii) positive reciprocity, and (iii) concerns for equality.

Result 5 *When decomposing trustworthiness, strategic motives appear as a major drive, comparable in size to positive reciprocity, and larger than concerns for equality.*

Figure 5: Determinants of trustworthiness

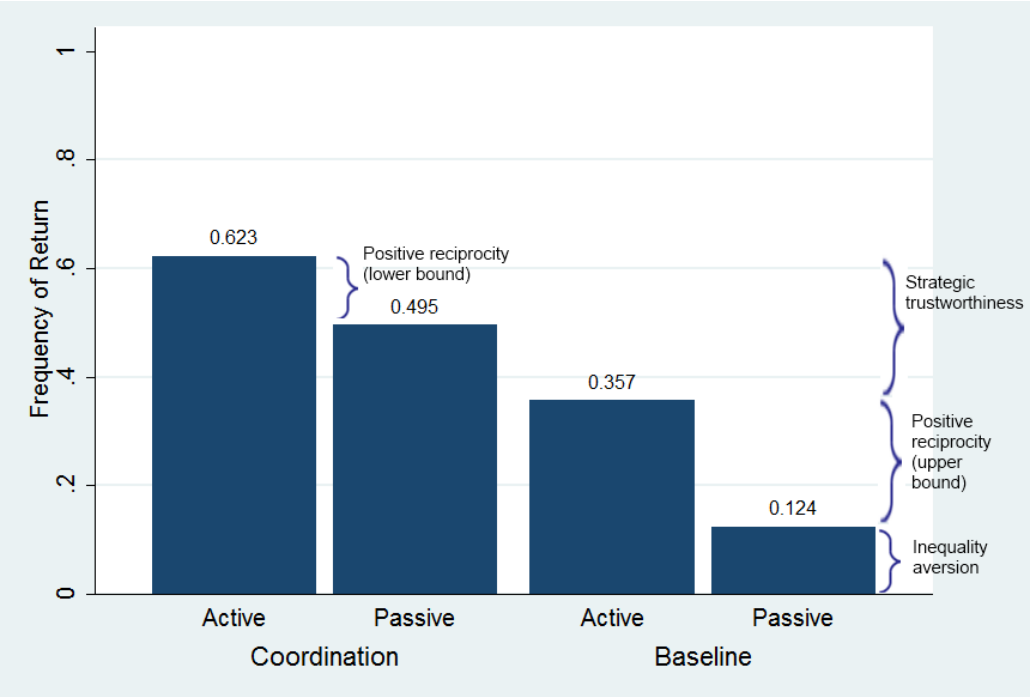


Table 7: Determinants of trustworthiness: All treatments

Dep. var.: *Return(1) Retain(0)*	Model (1)l	Model (2)
Coordination (d)	0.422***	0.419***
	(0.044)	(0.051)
Active (d)	0.296***	0.271***
	(0.069)	(0.078)
Active x Coordination (d)	-0.181	-0.191*
	(0.120)	(0.115)
Demographics	No	Yes
N.obs.	2276	2276
Pseudo-R2	0.088	0.109

Notes: Marginal effects from a probit regression: standard errors robust for clustering at the matching-group level (in parentheses).

The strategic component of trustworthiness can be isolated by contrasting Return frequencies in *Coordination-Active* and *Baseline-Active* (mean Return frequencies of 0.62 vs. 0.36, respectively, Figure 5). We attribute to concerns for equality the choice of Return in the *Baseline-Passive* treatment: the truster would earn 0 instead of 36 if none of the trustees returns (mean Return frequency is 0.12). Finally, positive reciprocity – the willingness to repay a kind and intentional action of trust – is present by design in all *Active* treatments but not in *Passive* treatments, in which a robot decides between Keep or Send. There are two ways to estimate the relative importance of reciprocal motives for trustworthiness. Both are given by the difference in trustworthiness between *Active* and *Passive* but one is the difference in the *Baseline* treatment (0.36 vs.0.12) and the other is the same difference in the *Coordination* treatment (0.62 vs. 0.50). The estimated impact of positive reciprocity is smaller in the latter comparison. Our interpretation of this difference in differences is that subjects can have more than one motive to be trustworthy: some of the subjects who would be trustworthy for reciprocal reasons in the *Coordination-Active* treatment, could be already driven to be trustworthy for the strategic motivation in the *Coordination-Passive* treatment. The strategic motivation 'captures' reciprocal subjects and make it seems as if these are fewer than they really are. In other words, we believe that the best estimate of reciprocity as a motive of trustworthiness is the difference between *Active* and *Passive* in the *Baseline* treatment rather than in the *Coordination* treatment, where the difference is underestimated. Table 7 provides additional evidence about these two estimates of reciprocal motives. Table 7 presents a series of probit regressions where the dependent variable is whether the trustee chose Return (1) or Retain (0) in each period. The *Coordination* dummy captures the strategic component of trustworthiness and is positive and highly significant in all estimates. The *Active* dummy is meant to capture positive reciprocity; while the coefficient is positive, it is noticeable that the marginal effect of positive reciprocity is smaller than the one of strategic trustworthiness. The crowding-out effect between strategic trustworthiness and positive reciprocity is captured by the interaction variable *Active* $\times$ *Coordination*. To sum up, the decomposition of trustworthiness in the *Coordination-Active* treatments yields the following quantitative results. The total of 62 percent of Return choices is attributed for 26 percentage points to strategic motives, 24 percentage points to positive reciprocity, and 12 percentage points to equality concerns. It is quite remarkable that in our multi-players trust game strategic trustworthiness is comparable in size to (if not larger than) positive reciprocity.

5 Discussion and Conclusions

We extend the standard conceptualization of trust as a dyadic interaction to a situation with multiple trustees. Our setting, the Collective Trust Game, introduces two novel elements: first, trustees' payoffs are interdependent and what trustees choose to do depends

on what they think other trustees will do; second, and because of that, trusters' decisions will also depend on what they believe is the predominant convention in their communities with regard to trustworthiness.

While this new setting fits standard definitions of trust, it also captures a new trust-warranting property absent from previous experiments. We introduce a coordination game among trustees to test whether even those who have no primary motive to act trustworthily still manage to do so if they expect that the other trustee is more likely to choose to be trustworthy than not (or if they believe that the other trustee will believe that they will be trustworthy and so on). Expectations about other trustees' behavior can thus lead self-interested payoff maximizers to be trustworthy. We report that strategic motives strongly affect both trust and trustworthiness. When decomposing trustworthiness, strategic motives appear as a major drive, comparable in size to positive reciprocity, and larger than concerns for equality. Moreover, higher levels of trustworthiness also led to higher levels of trust.

Our extension to multiple trustees is only one among many possible ways to generalize dyadic trust, and the relative contribution of each motivations reported in this study is of course linked to the specific parameters we used. These considerations, though, should not distract from the evidence of a large and qualitatively different component of trust and trustworthiness that standard experimental investigations cannot capture. Situations where trustworthiness by convention matters exhibit a peculiar patterns of behavior; trustworthiness by convention not only affects the level of trust and trustworthiness, but also shapes learning dynamics, the role of risk preferences, and the accuracy of beliefs. When comparing norms and conventions of trust across societies, what may vary could be the strategic motivation to trust. If that is the case, the Collective Trust Game could provide a better proxy to capture cross-cultural differences than simpler games.

Our results carry a more optimistic implication than the dyadic trust game experiment: communities in which trustworthiness is believed to be the predominant course of action, generate more of trustworthiness than one could gauge simply taking into account the trustworthiness generated by pro-social virtues. Trustworthiness has a positive externality – is "contagious", and its beneficial bacteria spread to those who, as self-interested maximisers, would not otherwise be exposed to it. One more reason why, as Hirschman (1984) once suggested, trust is a most peculiar resource that is depleted by not being used.

Acknowledgments The authors thank Tim Cason, Enrique Fatas, Bendikt Hermann, Giorgio Negroni, participants at the 2011 Florence Experimental Economics workshop, the ESA 2011 meetings in Chicago and Luxembourg, 2011 Workshop on Trust and Cultural Evolution in Valencia, and seminar participants at Purdue University and

Bologna University. We gratefully acknowledge the financial support from the ERC grant Strangers 241196. The usual disclaimer applies.

6 Appendix

Evolutionary stability of trust in the CTG

In this appendix, we show that the cooperative strategy (C) that prescribes to Send when in the role of truster, and to Return when in the role of trustee is not evolutionary stable in the *Baseline* variation of the Collective Trust Game, while it is evolutionary stable in the *Coordination* variation.

Baseline variation

Consider a large population of individuals, all of whom are "programmed" to play the cooperative strategy C. Suppose a small group of "mutants" – programmed to play strategy D – appear in this population. Let the share of mutants in the (post entry) population be $\varepsilon \in (0,1)$. In every period, sets of three individuals in this population are randomly formed, and roles are also assigned at random, so that each individual has the role of truster with probability $\frac{1}{3}$, and the role of trustee with probability $\frac{2}{3}$. In addition, for each individual the probability that each of the other 2 members of the set will play the mutant strategy D is ε and the probability that she will play the incumbent strategy C is $1-\varepsilon$. The expected payoff in a period in this bi-morphic population is thus the same as in a match with an individual who plays the mixed strategy $M_1 = \varepsilon D + (1\varepsilon)C$. The expected payoff to the cooperative strategy C is $u(C, M_1)$ and that of the mutant strategy $u(D, M_1)$.

According to Weibull (1997), a strategy x is *evolutionary stable* if for every strategy $y \neq x$ there exists some $\varepsilon_y \in (0,1)$ such that:

$$u[x, \varepsilon y + (1-\varepsilon)x] > u[y, \varepsilon y + (1-\varepsilon)x], \;\; \forall \varepsilon \in (0, \varepsilon_y)$$

The intuition is that evolutionary forces will select against a mutant strategy only if that strategy is less "fit" than the incumbent strategy, i.e. if in expectation it generates a lower payoff.

It is straightforward to show that, in the *Baseline* variant of the Collective Trust Game, the cooperative strategy C is not evolutionary stable. In fact, it can be invaded by a mutant strategy D_1, which prescribes to Send when in the role of truster, and to Retain when in the role of trustee.

The incumbent strategy C yields an expected payoff equal to:

$$u[C, \varepsilon D_1 + (1-\varepsilon)C] = \frac{1}{3}\left(1-\varepsilon^2\right)36 + \frac{2}{3}30 = 32 - 12\varepsilon^2$$

while the "mutant" strategy D_1 yields an expected payoff equal to:

$$u[D_1, \varepsilon D_1 + (1-\varepsilon)C] = \frac{1}{3}\left(1-\varepsilon^2\right)36 + \frac{2}{3}\left[\varepsilon 48 + (1-\varepsilon)\,30\right] = 32 - 12\varepsilon^2 + 12\varepsilon.$$

Hence,

$$u[C, \varepsilon D_1 + (1\varepsilon)C] < u[C, \varepsilon D_1 + (1\varepsilon)C], \ \ \forall \varepsilon > 0.$$

Along the same lines, one could show that none of the four possible pure strategies available in this game is evolutionary stable.[13]

Coordination variation

A similar reasoning shows that the cooperative strategy C is evolutionary stable, as it cannot be invaded neither by any of the three other possible pure strategies, nor by the mixed strategy S which prescribes to Send when in the role of a truster, and to Return with probability $\frac{1}{3}$ when in the role of a trustee.

The intuition is that any mutant strategy which prescribes to play Keep when in the role of a truster grants the truster a payoff of 20 units, which – for ε low enough – is lower than the expected payoff the truster gets by playing Send (which is at least equal to $36 - 36\varepsilon^2$ units). On the other hand, any strategy that prescribes to Retain when in the role of a trustee yields the trustee an expected payoff of $48\varepsilon + 20(1 - \varepsilon)$ when the truester chooses to Send, which – for ε low enough – is lower than the sure payoff of 30 the trustee would get by playing the cooperative strategy C.

[13]The exact computations are available from the authors, upon request.

Instructions

Treatments: Baseline-Active and [Baseline-Passive]

Welcome! This study is part of a research project of the University of Bologna and is financed by the European Commission.
You will earn money depending on your choices and the choices of the other participants. You will be paid in private at the end of today's study.
Please, turn off your mobile phone. From this moment on, no form of communication among participants is allowed. In case you have a question, please rise your hand and one of us will come to your desk to answer it.
Please, follow the instructions carefully. In this study there are three parts; I am about to read instructions for Part 1.

Instructions: Part 1

In this part, you have to choose among six different earnings options. Each option can produce either a high or a low earning. Please, look at the screen; for each option:

- the high earning is in the second column;
- the low earning is in the third column.

The high earning has a 50 percent probability to be realized and the low earning has a 50 percent probability to be realized.
What is your task? You have to choose your favorite option. Look at the screen: there is a button – I to VI – for each row. In order to choose, you have to press the button next to your favourite option. Please, touch the screen only with your fingers; pencils could damage the screen.
How are your earnings computed?

- at the end of today's study, two participants in this room will be selected at random; only the selected participants will receive a payment for this part. The payment can be either high or low;
- there will be a urn containing ten balls: 5 orange and 5 withe balls;
- if an orange ball is drawn from the urn, the selected participants will get the high payment for the selected option;
- if a withe ball is drawn from the urn, the selected participants will get the low payment for the selected option.

Let consider an example with no consequences for your final earnings. Please, press the button V. You earn Euros 37.5 if an orange ball is drawn from the urn, while you earn Euros 7.5 if a withe ball is drawn from the urn. *[Have you pressed the button?]* A box to CONFIRM or CHANGE your choice has appeared on the screen; please, press CHANGE.

Now you can change your choice. *[OK]* Now press the button III; you earn Euros 27.5 if an orange ball is drown from the urn, while you earn Euros 12.5 if a withe ball is drown from the urn. Please, press CONFIRM. *[The choice cannot be changed anymore. Please, everyone press CONFIRM. Is everything clear? If there are no questions, we can start with Part 1.]*

Instructions: Part 2

In the present and subsequent parts, your earnings are expressed in tokens: tokens will be converted in Euros at the rate of 1 Euro for 40 tokens.
In this part, people in this room are randomly divided into groups of three; nobody can know the identity of the other members of the group.
Three situations will be presented in turn. What is a situation? Look at the screen, you can see an example of a situation.*[Can you see six figures on the screen? OK]* Each figure is divided in three slices; the red slice indicates your earnings, while the black and the withe slices represent the earnings of the other people in your group.
For instance, in the top-left figure, your earnings are equal to 160 tokens, as well as the earnings of the other two members of your group. Let consider another example; if you choose the bottom-right figure, you earn 160 tokens, while one person in your group earns 340 tokens and the other earns 130 tokens. *[Is everything clear?]*
As you can see, different figures can have different dimensions. The sum of the earnings of each member of the group are displayed below each figure. *[Is there any question?]*
What is your task? You have to choose one of the six figures; in order to choose, you have to press the figure you prefer the most. As an example without consequences for your earning, press the top-center figure. A box to "CONFIRM" or "CHANGE" your choice has appeared on the screen; the chosen figure is highlighted by a withe box. Please, press CONFIRM. *[Your choice cannot be changed now. Is there any question?]*

How are your earnings computed? Every person in your group will make a choice for each situation. Among all the choices made within your group, only one randomly chosen choice will be implemented; the implemented choice can be your choice or the choice by another member of your group.
What if your choice is chosen at random? Your choice will determine your earnings and the earnings of the other members of your group.
What if the choice of another person in your group is chosen at random? It can be the case that your earnings are different from the one you chose. In this case, your earning depends on the choice made by the selected person and the color you have been assigned at random: either withe or black. The choice will be randomly selected at the end of this part; therefore, you have to pay attention to all of your choices. *[Is everything clear?]*
Before starting, please answer a few questions.

Instructions: Part 3

As in the previous part, people in this room are divided at random in groups of three people; nobody can know the identity of the other members of the group.
There are two different roles in each group: role A and role B. In each group, a person has the role A and two have the role B. Roles are randomly assigned by the computer.

What is the task for a role A person? Look at the screen, A has to decide between PASS and KEEP.

- if the person A decides to KEEP, everyone in the group earns 20 tokens. In this case, people playing as B do not have to take any decision;
- if the person A decides to PASS, earns either 0 or 36 depending on the choices made by people playing as B in the group.

As an example with no consequences for your earnings please, press the button PASS. You can now decide whether to CONFIRM or CHANGE your choice; please, press CONFIRM. *[The choice cannot be changed anymore. Please, everyone press CONFIRM.]*

[The computer will draw a ball from an urn containing blue and yellow balls:

- In case the ball is yellow, everyone in the group earns 20 tokens. In this case, people playing as B do not have to take any decision;
- In case the ball is blue, A earns either 0 or 36 depending on the choices made by people playing as B in the group.

The person playing as role A has no choice to make.]

What is the task for B if A decides to PASS [if the ball is blue]? Look at the screen, the two participants with role B have to simultaneously choose between GIVE and KEEP. If A decides to PASS [If the ball is blue], how are the earnings computed?

- if both Bs decide to KEEP, Bs earn 48 tokens each and A earns 0 tokens;
- if both Bs decide to GIVE, Bs earn 30 tokens each and A earns 36 tokens;
- if one B decides to GIVE and the other decides to KEEP, Bs earn 30 tokens each and A earns 36 tokens.

Please remember that everyone in the group earns 20 tokens if A decides to KEEP [if the ball is yellow].
Let consider an example with no consequences for your earnings. In this example, you have been assigned to role B. Please, press KEEP and then CONFIRM. *[The choice*

cannot be changed anymore. Please, everyone press CONFIRM]
In the following screen, you can see the final earnings for the group. In the present example, you have been assigned role B and have decided to KEEP, while the other person who have been assigned to role B decided to GIVE. Both you and the other person with the role B earn 30 tokens and A earns 36 tokens.
In this part there are 30 rounds with the same rules. In the upper-left part of the screen you can see the number of the current round. At the beginning of every round, new groups of three people are formed at random.
To sum up, in every round:

- A has to decide between PASS and KEEP;
- if A decides to KEEP [if the ball is yellow], the round ends and everyone earns 20 tokens;
- if A decides to PASS [if the ball is blue], the two people with role B have to choose between GIVE and KEEP and earnings are computed as explained above.

Earnings cumulate from round to round. The composition of the urn has been decided in advance and does not depend on the choices people in this room will make.

[Is everything clear?] Before starting, please answer a few questions.

References

Bacharach, M. and D. Gambetta (2001). Trust in signs. *Trust in society 2*, 148–184.

Bacharach, M., G. Guerra, and D. Zizzo (2007). The self-fulfilling property of trust: An experimental study. *Theory and Decision 63*(4), 349–388.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. *Games and Economic Behavior 10*(1), 122–142.

Burks, S., J. Carpenter, and E. Verhoogen (2003). Playing both roles in the trust game. *Journal of Economic Behavior & Organization 51*(2), 195–216.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Cason, T. and V. Mui (2007). Communication and coordination in the laboratory collective resistance game. *Experimental Economics 10*(3), 251–267.

Cassar, A. and M. Rigdon (2011). Trust and trustworthiness in networked exchange. *Games and Economic Behavior 71*(2), 282–303.

Coleman, J. (1990). *Foundations of social theory*. The Belknap Press of Harvard University Press (Cambridge, Massachussets.).

Cox, J. (2004). How to identify trust and reciprocity. *Games and Economic Behavior 46*(2), 260–281.

Croson, R. T. (2000). Thinking like a game theorist: factors affecting the frequency of equilibrium play. *Journal of Economic Behavior & Organization 41*(3), 299 – 314.

Dasgupta, P. (1988). Trust as a commodity. In D. Gambetta (Ed.), *Trust. Making and Breaking Cooperative Relations.*, pp. 49–72. New York: Basil Blackwell.

Eckel, C. and R. Wilson (2004). Is trust a risky decision? *Journal of Economic Behavior & Organization 55*(4), 447–465.

Eckel, C. C. and P. J. Grossman (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior 23*(4), 281 – 295.

Eckel, C. C. and P. J. Grossman (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization 68*(1), 1 – 17.

Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association 7*(2-3), 235–266.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Greiner, B. (2004). The online recruitment system orsee 2.0-a guide for the organization of experiments in economics. University of Cologne, Working paper series in economics, 10.

Hirschman, A. O. (1984). Against parsimony: three easy ways of complicating some categories of economic discourse. *American Economic Review Papers and Proceedings 74*, 88–96.

Houser, D., D. Schunk, and J. Winter (2010). Distinguishing trust from risk: An anatomy of the investment game. *Journal of Economic Behavior & Organization 74*(1-2), 72–81.

Jacquemet, M. (1996). *Credibility in court: Communicative practices in the Camorra trials*, Volume 14. Cambridge University Press.

Johnson, N. and A. Mislin (2011). Trust games: A meta-analysis. *Journal of Economic Psychology 32*(5), 865–889.

McCabe, K., M. Rigdon, and V. Smith (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization 52*(2), 267–275.

McEvily, B., R. Weber, C. Bicchieri, and V. Ho (2006). Can groups be trusted? an experimental study of trust in collective entities. In A. Z. Reinhard Bachmann (Ed.), *Handbook of trust research*, pp. 52–67. Cheltenham, UK: Edward Elgar.

Mittone, L. and M. Ploner (2009). Social effects in a multi-agent investment game. an experimental analysis. CEEL Working Papers.

Regner, T. and G. Riener (2011). Motivational cherry picking. Jena Economic Research Papers.

Rousseau, D., S. Sitkin, R. Burt, and C. Camerer (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review 23*(3), 393–404.

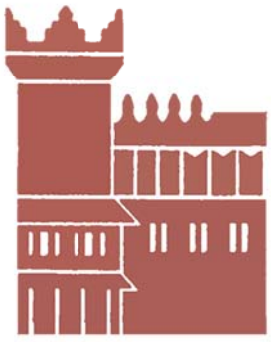
Sapienza, P., A. Toldra, and L. Zingales (2007). Understanding trust.

Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural Paraguay. *Journal of Economic Behavior & Organization 62*(2), 272–292.

Weibull, J. (1997). *Evolutionary game theory*. The MIT press.

Weingast, B. (1997). The political foundations of democracy and the rule of law. *American Political Science Review*, 245–263.

Young, H. P. (2008). Social norms. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Basingstoke: Palgrave Macmillan.



Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it