Collegio Carlo Alberto

Intermediaries in Trust: An Experimental Study on Incentives and Norms

Giangiacomo Bravo
Flaminio Squazzoni
Károly Takács

No. 240
February 2012

Carlo Alberto Notebooks

www.carloalberto.org/working_papers

Intermediaries in Trust: An Experimental Study on Incentives and Norms

Giangiacomo Bravo[*], Flaminio Squazzoni[†], Károly Takács[‡]

February 2, 2012

Abstract

Trust situations involve a certain amount of risk for trustors that trustees can abuse. Knowing this, rational individuals might not place trust in others. In many real situations, intermediaries exist that play a crucial role helping trustors and trustees to cooperate. The question is how intermediaries can be motivated to accomplish this important task. We have investigated this by performing various experimental tests on a modified version of the investment game, where we introduced intermediaries who rated the behavior of trustees for the benefit of trustors. We manipulated incentive schemes and tested various role structures. We found that intrinsic motivations can increase the positive impact of the intermediaries on cooperation between trustors and trustees, even more than any material incentive, provided that there is room for indirect reciprocity strategies. Our results show the importance of intrinsic motivations of individuals and circumscribe the power and generalizability of material incentives for cooperation.

Keywords: trust; intermediaries; cooperation; investment game; incentive schemes; indirect reciprocity.

1 Introduction

A trust relation is an exchange where at least two parties are present, a trustor and a trustee, which involves a certain amount of risk for the former. If the trustor decides to place trust, the trustee can honor or abuse it. If honoring trust is costly, and this is the rule at least in one-shot exchanges, the trustee will have no rational incentive to be trustworthy. Knowing this, it is likely that the trustor does not even enter in the exchange (see Coleman 1990, Chap. 5).

Given that no rational solution exists to the problem of trust in one-shot exchanges, sociologists and economists have racked their brain to understand how trust can be established in such situations. One of the most appealing explanations indicated that social and economic exchanges mostly take place in social contexts and are embedded in social life

[*]University of Torino and Collegio Carlo Alberto <giangiacomo.bravo@unito.it>.

[†]Department of Social Sciences, University of Brescia <squazzon@eco.unibs.it>.

[‡]Institute of Sociology and Social Policy, Corvinus University of Budapest <karoly.takacs@unicorvinus.hu>.

(e.g., Barrera 2005; Cook and Hardin 2001). This means that people's behavior, even in case of an exchange between strangers, does not occur in a vacuum as certain habits, practices and roles evolved in social life that mediate between people. For instance, exchanges are frequently observed by bystanders, intermediaries might act as advisories and mediators, and record-keeping or gossip allow individuals to spread information about potential partners in many social and economic spheres (e.g., Burt and Knez 1995; Coleman 1990, 184–185).

Recent experimental results showed that people trust and cooperate more if they think that someone else can observe their behavior (Bateson et al. 2006; Bohnet and Frey 1999; Cason and Mui 1997; Haley and Fessler 2005; Rigdon et al. 2009). The presence of intermediaries can lead to a higher level of trust in conditions of information asymmetry between trustors and trustees (Boero et al. 2009a). As intermediaries are in a crucial position to evaluate others' behavior and are often called to express their opinion for the benefit of trustors or trustees, it is of paramount importance to understand under which conditions intermediaries' opinion is efficiently used by others and how intermediaries can be motivated to provide informative evaluations that contribute to increase cooperation. For instance, in many important social and economic interactions—such as in the trust relationships between shareholders and managers in a company, between submitting authors and journal editors in science or between customers and producers of cultural products—intermediaries exist that express opinions which are crucial to create trust and favor cooperation. Sometimes, they do so voluntarily without receiving any material payoffs, such as referees in scientific journals. In other cases, intermediaries are financially motivated professionals, such as stock brokers or literary and art critics. It seems that there are certain spheres of social life where social and historical processes helped to institutionalize the social function of intermediaries through material incentives, others where informal or voluntary norms are predominant.

The aim of this paper is to look at these intermediaries in trust. We did so experimentally, by creating a series of experiments where individuals were called to play an investment game where trust was crucial to ensure cooperation between trustors and trustees. We manipulated incentives and roles of intermediaries to observe their impact on cooperation. The rest of the paper is organized as follows. Section 2 presents the theoretical background and the research hypotheses that guided our experiments. We will discuss simple and more complex trust relationships, by focusing on the crucial role of intermediaries. Section 3 and 4 illustrate our methods, report the experimental design and present the results while Section 5 discusses the results.

2 Research background and hypotheses

2.1 Simple trust relations

A typical formalization of a trust relation is the *Trust game* (Camerer and Weigelt 1988; Dasgupta 1988). In this game, Player A is called to decide whether to place trust on Player B. If he/she decides not to place trust, both players receive the "punishment" payoff P and the interaction does not occur. If he/she places trust, Player B can choose to honor or abuse it. If Player B abuses trust, he/she receives the "temptation" payoff T while Player A receives the "sucker" payoff S. If Player B honors it, both players receive the "reward"

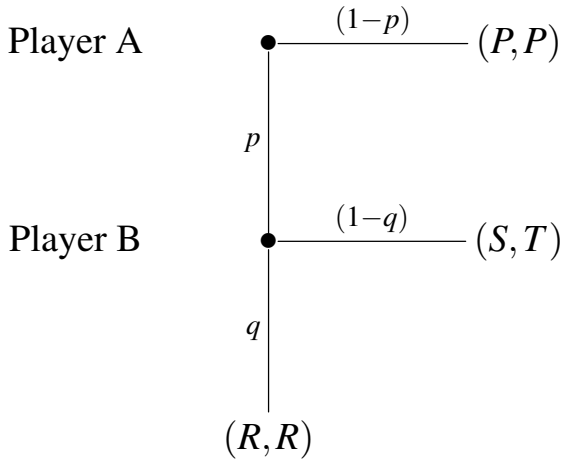

Figure 1: The Trust game

payoff R. The relation $T > R > P > S$ must hold to speak of a Trust game (Fig. 1).

In Figure 1, the parameter $p \in [0,1]$ represents the probability that Player A places trust and can be interpreted as a measure of his/her trust in B's cooperation. The parameter $q \in [0,1]$ represents the probability that Player B cooperates and can be interpreted as his/her trustworthiness. Despite a formal dominant strategy that predicts an expected absence of trust, if Player A believes that B will be trustworthy with a certain probability q, it is in his/her interest to place trust as long as $qR+(1-q)S > P \Rightarrow q > (P-S)/(R-S)$, which represents a measure of the risk borne by the trustor if he/she places trust (Bravo and Tamburino 2008; Buskens and Raub 2008).

On the other hand, it is likely that trust relations do not always imply all-none choices, such as to trust or not. The social and economic life is full of examples where trustors can decide how much trust to place in trustees. This is what the *Investment game*, first presented by Berg et al. (1995), looks at. The rules of the game are simple. First, Player A (the trustor) receives an initial endowment of d_A points, with a fixed exchange rate in real money. Player A has to decide the amount I between 0 and d_A to send to Player B (the trustee), keeping for him/herself the part $(d_A - I)$. The amount sent by A is multiplied by $m > 1$ and sent to the trustee, in addition to his/her own endowment d_B,[1]. Then B decided the share of the amount received (plus his/her endowment) to return to A. As before, the amount R returned by B can be any integer between 0 and $(d_B + mI)$. The amount $(d_B + mI - R)$ not returned represents B's profit, while R is summed to the part kept by A to form his/her final profit, which is calculated as $(d_A - I + R)$ (Fig. 2).

In the Investment game, A players have a serious interest in investing as long as the expected returns are higher than their investments, i.e, if $R = q(d_B + mI) > I$, where q is the proportion returned by B. This means that, if $d_B = 0$, $q > 1/m$ should hold for A investment to be profitable. Therefore, the investment will be d_A if A believes that B will return at least $1/m$ of the amount received. Otherwise, the investment will be zero. If $d_B > 0$, the relation between the amount that A should rationally invest and q is more complex. Generally, the higher is d_B, the lower the proportion of returns needed to make

[1]Note that in some Investments games, including Berg et al. (1995), B players had no endowment, i.e., $d_B = 0$. Johnson and Mislin (2011) found that, in experiments where $d_B > 0$, trustors' investments tend to be lower.

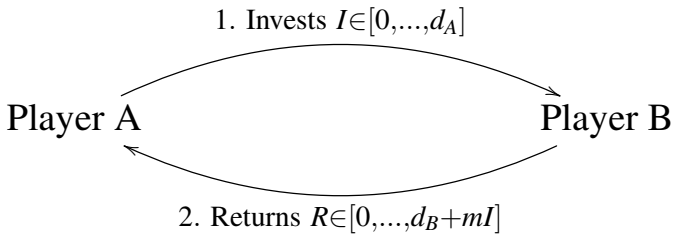

Figure 2: The Investment game. A earnings $= d_A - I + R$. B earnings $= d_B + mI - R$.

A investments profitable is. However, as the rational perspective of the trustee is not to return any proportion received and the trustor knows this, the theoretical prediction is that rational trustors should not invest anything in this game.

The dynamics of the Trust and the Investment games have been extensively explored experimentally. Despite theoretical predictions, in experiments based on the Trust game and played anonymously, about half of A players chose to place trust, while most B players were trustworthy (e.g., McCabe and Smith 2000; Snijders and Kerens 1998). The same is true for the Investment game. In a recent review of 162 replications of the Investment game, Johnson and Mislin (2011) calculated that, on average, A players invested about 50% (range 22–89%) of their endowment and B players returned 37% (range 11–81%) of the amount in their possession.

These experiments showed that various factors can affect trust decisions by trustors, including the frame of the game, subjects' age and the country where the game is played (see Johnson and Mislin 2011). However, the crucial issue from a theoretical point of view is to understand the mechanisms through which trustors can estimate the trustworthiness of trustees. Following Coleman (1990, 191), there are three possible sources of information that can act as trust carriers for trustors: (i) the direct knowledge on past behavior of the trustee; (ii) the knowledge obtained from a third party that has positions and interests similar to the trustor; (iii) the knowledge obtained from a third party holding different positions and interests, which are not linked with the trustor.

It is widely acknowledged that knowing past behavior of others can increase cooperation in this type of games.[2] In this case, the trustee has a rational incentive to acquire a good reputation (i.e. a reputation of being trustworthy) and this can motivate the trustor to increase his/her investment (Bravo and Tamburino 2008). This is confirmed by numerous experiments where subjects repeatedly played with the same partner. For instance, Cochard et al. (2004) and Barrera and Buskens (2009) found that the "dyadic embeddedness" of players increased investments and returns in a repeated investment game. More generally, under this condition, subjected trusted more as long as trust was honored, but trust rapidly declined after any abuse (see Buskens and Raub 2008). This means that a process of learning took place as subjects exploited the repetition of the interaction to understand whether their opponents were trustworthy.However, being the interaction dyadically embedded, reciprocity led to a rapid decline of cooperation once trust was abused.

[2]Note that both the Trust and the Investment game have the form of a social dilemma where both players can achieve higher results by cooperating.

In this study, we will concentrate on the second and third points made by Coleman. As we knew that other important sources of social information could be used to predict others' behavior, such as signals, social tags or stereotypes (e.g., Kim 2010; Hales 2000), we designed our experiments to intentionally exclude their effect.

2.2 Third-party trust relations

While dyadic embeddedness is important to explain trust and cooperation in situations involving stable relationships between two actors, in modern societies trust is often mediated by agents who facilitate the exchange between trustors and trustees when they cannot rely on past experience (see Coleman 1990, Chap 8; Cook et al. 2005; Hardin 2004). Important empirical examples of this have been found in the development of trust between suppliers and customers in a variety of situations, such as in electronic markets on the web (Bailey and Bakos 1997; Palmer et al. 2000) and in the U.S. venture capital market (Sorenson and Stuart 2001).

Following Coleman's argument, it is crucial to understand the role of trust intermediaries who have positions and interests either analogous or different from the trustors. When positions and interests are aligned, trustors are expected to seriously consider the opinion of the intermediaries so that their decisions will reflect reputational information available. Coleman (1990, 192) argued that this type of information may lead to homogeneous behavior as decisions to place trust is likely to reflect the past. For instance, Resnick and Zeckhauser (2002) examined exchanges on the eBay platform and found that the existing rating system tended to increase exchanges for sellers who achieved a good reputation, so reproducing trusting behavior towards the same sellers. In a theoretical study, Bravo and Tamburino (2008) showed that a system based on the knowledge of trustees' past behavior could sustain trust and a consistent degree of trustworthiness provided that information was sufficiently reliable.

Numerous experiments studied similar situations. Buskens et al. (2010) designed a simple Trust game played in triads composed by two trustors and one trustee. They found that cooperation significantly increased when trustors were allowed to exchange information about past behavior of trustees. This result is consistent with other experiments where various interaction structures were tested, which showed that, once information about past behavior was provided, people were willing to "invest"—i.e., to bear a cost in view of long term benefits—to acquire a good reputation (e.g., Fehr and Fischbacher 2003; Seinen and Schram 2006).

Keser (2003) designed an Investment game where subjects interacted repeatedly and trustors were allowed to rate the behavior of their opponents. Couples changed each round, but past ratings achieved by trustees were available to the new trustors before their investment decision. Her results showed that the rating system significantly increased cooperation, even if the introduction of the reputational opportunity in the game had a stronger effect on the proportion of the received amount returned by trustees (+41.5% in the "reputation" treatment in comparison with the baseline game) than that on the proportion of the endowment invested by A players (+31.7%). To sum up, Keser's rating system affected more trustworthiness than trust.

Boero et al. (2009b) extended Keser's results by introducing a two-way system, where both trustors and trustees were under rating. They showed that trustors under rating in-

creased their investment as trustees did with returns in Keser's experiment. However, although reputation building investments were rational for trustees as they were trustworthiness signals expected to increase trustors' investment, this was not true for trustors (who did not need to be trusted by trustees). Furthermore, these authors showed that returns even increased when any rational incentive for reputation building investments was ruled out by the experimenters, i.e., when past ratings of trustees were shown to trustors only *after* their investment decision. This suggest that, far from being only a matter of rational reputation building investment, individuals also look at social approval as an intrinsic motivation (see also Bateson et al. 2006; Haley and Fessler 2005).

The situation is different when the interests of intermediaries and trustors are not aligned, as the former are supposed to offer more "independent evidence" in support of trusting decisions, which "will be most likely to lead to a correct assessment" of trustees' reliability (Coleman 1990, 191). However, this strongly depends on the motivations behind intermediaries' actions. For instance, intermediaries whose interests are aligned with the trustees may be tempted to cheat trustors by offering unfair advices, as happened in the 2001 Enron scandal where Arthur Andersen, which was supposed to audit the company, did not alert the shareholders of clear credit risks because of a conflict of interests due to the high consulting fees offered by Enron (Swedberg 2005).

On the other hand, intermediaries who are not directly involved in the exchange might lack serious motivations for their behavior. Indeed, as they bear the cost without receiving any benefit, they might refuse to invest the time and energy needed to express a thoughtful evaluation. Without clear incentives for the intermediary, unreliable evaluations can even determine a reduction of trust in the long run.

To consider all these combinations, we suggest to distinguish four typical cases.

i. Intermediaries have material incentives that are aligned with the trustors. Here, intermediaries have a direct interest to provide clear and reliable advices to the trustors. This allows trustors to be confident about the reliability of these advices and use them to decide what to do. For instance, suppose that a merchant broker is trying to sell a bond of his/her bank to a costumer who knows that the broker is paid only by a share of his/her future benefit from the bond. The bond will appear trustworthy as it is in the interest of the broker to carefully fulfill the interest of the customer.

ii. Intermediaries have material incentives aligned with the trustees. Here, intermediaries are in a potential conflict of interest as they may be tempted to cheat trustors by providing advices leading to higher benefits for trustees. For instance, suppose you are director of a department carefully evaluating a set of competitors for a job position and a colleague of yours strongly recommends a member of his/her family who is competing. Knowing this, you (the trustor) will no longer consider your colleague (the intermediary) as a reliable source of information and will consequently be less willing to accept the candidate as a serious competitor for the position.

iii. Intermediaries have material incentives independent of the interests of both trustors and trustees. This implies that a third party not involved in the exchange provides intermediaries with a "wage" to motivate their work. For instance, this is the case of literary or art critics, who are professionally engaged by independent organizations (e.g., newspapers or magazines) to assess specific cultural or artistic products, such as a book or an exhibition, with direct material interests neither with artists not with customers. In this case, previous studies showed that incentives work only if they are

of a sufficient magnitude to seriously motivate intermediaries (see Heyman and Ariely 2004).

iv. Intermediaries do not respond to specific material incentives but are intrinsically motivated to carefully accomplish their task. If trustors acknowledge intermediaries' motivations, the intermediaries' advices will be seriously considered to decide whether to place trust on the trustees. This is, for instance, the case of peer review in scientific journals, where referees are supposed to carefully evaluate authors' submissions without receiving any material benefit in exchange.

On the other hand, it is argued that intrinsic and material incentives may even conflict. A recent theory, called *Motivation crowding theory*, has been elaborated that accounts for a broad range of empirical phenomena where external interventions, such as monetary incentives or fines, might undermine intrinsic pro-social motivations so as to dominate the traditional relative price effect (Frey and Jegen 2001). One of the most famous examples was the unintended effect of monetary incentives on blood donations illustrated by Titmuss (1971), where the introduction of economic incentives to increase donations degraded the donors' willingness to contribute by eroding the moral foundation of the donors' decision (e.g., Mellström and Johannesson 2008). The same was also found in a field study on a group of day-care centers in Israel, where monetary fines for parents who were late picking up their children were imposed, with the result that parents doubled the delay and still continued to do it 12 weeks after the revocation of the fine (Gneezy and Rustichini 2000).

More generally, Ariely et al. (2007) performed a mixed study that combined laboratory experiment and field observation to understand the relationship between incentives and charitable donations. They found that extrinsic monetary incentives had a detrimental effect on pro-social motivations and determined a decrease of donations. Meier (2007) found the same result in a field experiment on charitable giving in organizations: when subsidies were introduced to promote the willingness of the company employees to contribute, donations drastically decreased. More recently, a social experiment in the catchment areas of two large Norwegian hospitals by Holmaas et al. (2010) arrived at the same conclusion: by imposing fines on municipalities (the hospital owners) to reduce the length of stay, the length increased in the long-run as fines transformed a moral problem into a cost/benefit calculus.

Fehr and Falk (2002) argued that, by viewing individuals as mere utility maximizers rationally responding to material incentives, we are even limiting our progress in understanding incentives. They reviewed the experimental behavioral literature and showed that material incentives could backfire and reduce individual efforts and achievements in various economic exchange situations, e.g., from team collaboration to shareholder/manager relationships. They argued that social incentive provision could be viewed as a powerful alternative to monetary incentives in motivating cooperation.

Note also that strong non-material incentives for intermediaries to cooperate with trustors comes from indirect reciprocity motives. Indirect reciprocity was introduced by Alexander (1987), even if the idea was first formulated by classical economic anthropologists (e.g., Malinowski 1922; Mauss 1954; Polanyi 1968). In sociology, the same idea has been sometimes popularized under the concept of "generalized exchange" (Takahashi 2000). The idea is that indirect reciprocity create a chain of benefit giving, or "helping", behavior among $n > 2$ agents. Here, when A_1 helps A_2, the latter does not directly re-

ciprocate but can help a third agent A_3. In turn, A_3 will help A_4 and so on until A_n helps A_1, so closing the circle. Indirect reciprocity can be socially efficient to overcome certain limitations of direct reciprocity, i.e., the necessity of repeated interaction between the same subjects. Nowak and Sigmund (1998a,b) showed that through "image scoring", individual selection favored the evolution of cooperative strategies even if the exchanges between the interacting subjects were neither repeated nor simultaneous. By image scoring, the authors simply denoted the attribution to each player of a score, which was directly proportional to his/her cooperative past behavior. If players could modify their behavior as a function of the image of their opponents, cooperating with others was effective as a reputation building investment.

Several experiments explored the dynamics of indirect reciprocity. For instance, Milinski et al. (2002) presented an experiment based on the alternation of Public good and Indirect reciprocity games modeling common-pool resource management problems. Their study showed that the need for maintaining a good reputation for the Indirect reciprocity game led to higher cooperation even in the Public good game. In another experiment, Seinen and Schram (2006) showed that indirect reciprocity strongly motivated the cooperative behavior of subjects. In this study, subjects played a repeated Helping game in groups of 28 for at least 90 rounds. In each period, they were coupled and played the game in pairs either as "donor" or as "recipient", where the donor had to decide whether to help the recipient at a personal cost or to "pass", avoiding the cost, but also giving no benefit to the recipient (note that the donor cost was less than the recipient benefit). These authors showed that introducing the possibility of knowing the recipient actions of the past 6 rounds drastically increased the helping frequency in the game. They explained this as a clear effect of reputation-based strategies on individual behavior.

Moreover, the importance of indirect reciprocity for human behavior can explain the astonishing weight maintained by gossip also in contemporary societies where other, more reliable, sources of information are easily available (Nowak and Highfield 2011). In a series of experiments, Sommerfeld et al. (2007, 2008) examined the crucial role of gossip in solving collective action problems. They found that gossip sustained reciprocity between cooperators, so favoring a reduction of free riding behaviors. They also found that subjects frequently spread information about past behavior of others even if they did not derive any material interest from this. On the other hand, gossip was effective only because subjects were motivated to spread information about others' cooperative behavior. Indeed, being involved in a collective action dilemma, gossipers estimated to receive valuable information for their advices in turn.

These results allow us to understand that, in any typical third-party trust relationship, the intermediaries in trust, who might actively spread information about the trustworthiness of trustees, can have sufficient intrinsic motivation only if their role is not confined to broadcasting others' reputation. In other words, even if intermediaries do not directly participate in the exchange under consideration, the fact that they could potentially find themselves in the position of trustors could provide a sufficient motivation to provide a reliable evaluation of trustee behavior. If not, no indirect reciprocity can take place and it is difficult that trustors can trust intermediaries' evaluations.

2.3 Research hypotheses

Previous studies showed that intermediaries not directly involved in an exchange between a trustor and a trustee should increase cooperation as they favor trust from the trustors and trustworthiness from the trustees (Coleman 1990, Chap. 8). On the other hand, it must be recognized that the interplay of material incentives, social norms and reciprocity can create a complex social interaction where the effect of intermediaries is neither necessarily robust nor even positive for cooperation. Therefore, disentangling these interaction aspects is essential to understand implications of intermediary behavior for trust and cooperation. To look at this problem, we formulated five research hypotheses to be experimentally tested. We explored each hypothesis by comparing it with a baseline situation where trustors and trustees interacted without any intermediaries.

Hypothesis 1 *If intermediaries respond to material incentives that are aligned with the trustors' interest, trustors and trustees will cooperate more.*

If the payoffs of intermediaries and trustors are aligned, there are clear incentives for the former to carefully evaluate the trustees' behavior for the benefit of the latter. This transform any trust relations in a typical principal-agent model. In this case, the rational choice literature dictates that monetary incentives are crucial to motivate trustees to act on investors' behalf, by guaranteeing that the self-interest of the former coincides with the objectives of the latter (Laffont and Martimort 2002). Therefore, we expect that trustors can rationally follow intermediary advices by increasing their investment with "good" trustees and decreasing it with "bad" ones. Knowing this, trustees can be induced to be more trustworthy, increasing cooperation in the system.

Hypothesis 2 *If intermediaries respond to material incentives that are aligned with the trustees' interests, trustors and trustees will cooperate less.*

If the payoffs of intermediaries are aligned with trustees, intermediaries have an incentive to cheat trustors by providing advices intended to favor trustees. Indeed, they are in a conflict of interest as they are expected to generate a benefit to the trustors, but their payoffs depend on the benefit of the trustees that they should evaluate. Consequently, trustors will view intermediary advices as unreliable and tend to reduce investments. This, in turn, will lead to lower returns for trustees and will erode cooperation.

Hypothesis 3 *If intermediaries respond to material incentives that are fixed and independent both from the trustors' and the trustees' interests, cooperation will increase significantly but less than in H1.*

In this case, although in a position of independence of judgment, intermediary incentives are less intelligible for trustors and trustees as they do not adapt to interaction outcomes. Moreover, following *Motivation crowding theory*, material incentives will rule out those intrinsic incentives for fair and careful evaluations. This makes intermediary incentives poorly credible for both trustors and trustees. As a result, we expect that cooperation will be lower than in the H1 case and, possibly, even close to the baseline, i.e., without intermediaries.

Hypothesis 4 *If any material incentive is ruled out from the picture, intermediaries can find intrinsic motivations and ensure high cooperation as they will be perceived as 'impartial' third-parties by the other players.*

Without having material interests in the exchange, intermediaries can perceive themselves as impartial judges of the exchange and provide informative opinion to trustors. Knowing this, trustors can place trust in intermediary opinions and trustees can be motivated to increase their trustworthiness, so determining more cooperation compared to incentive schemes. The absence of material incentives can transform the interaction in a moral problem, exploiting the impact of the disinterestedness of intermediaries on the other figures.

Hypothesis 5 *As fixed intermediary roles do not permit that indirect reciprocity takes place among subjects, the effect of intermediaries on cooperation in all experimental conditions will be weak. This effect will be even more pronounced when material incentives are not present.*

No indirect reciprocity strategy can work in a system where interaction roles are fixed, as the benefit of keeping reliable standards of evaluation for intermediaries does not have a positive impact on the future payoffs when the subject will play as trustor. In this case, any motivation of intermediaries to provide careful evaluations of trustees' behavior will disappear, especially when no clear material incentives are present.

3 Methods

To check our hypotheses, we built two experiments based on the Investment game described above, introducing intermediaries not directly involved in the transaction but asked to rate trustees' behavior. Evaluations were then transmitted to trustors before their investment decisions. In the first experiment, game roles (i.e., trustor, trustee and intermediary) alternated regularly throughout the game, while in the second they were fixed. In both experiments, incentives for intermediaries systematically varied across treatments. Sections 3.1 and 3.2 provide a detailed description of the two experiments.

3.1 Experiment 1 (alternating roles)

A total of 136 subjects (50% females) participated in the experiment held at the Faculty of Economics of the University of Brescia in the late 2010. Participants were students recruited across the university using the online system ORSEE (Greiner 2004). All interactions were anonymous and took place through a computer network equipped with the experimental software z-Tree (Fischbacher 2007).

Twenty-eighth subjects participated in a baseline repeated Investment game (hereafter *Baseline*) set using the following parameters: $d_A = d_B = 10$ monetary units (MU), $m = 3$. Each MU was worth 2.5 Euro Cents and subjects were payed in cash immediately after the end of the experiment. The game was repeated 30 times with couples that were randomly reshuffled after each period. Players' roles regularly alternated throughout the game. This means that each subject played exactly 15 times as A and 15 times as B.

All the other treatments, each played by 27 subjects, introduced a third player into the game (player C) in the role of intermediary. If selected as C, subjects were informed of the amount received and returned by the B players in the previous period. Then, C players were asked to rate B players' behavior as "negative", "neutral" or "positive". Reviews were displayed to A players before the subsequent investment decisions. As before roles alternated throughout the game, with subjects that played exactly 10 times in each role.

Once C player were introduced, we varied the incentive schemes offered to them. In the *No incentive* treatment, subjects did not receive any reward for reviewing. When applied to this interaction scheme, the incentive-based rational choice perspective predicts that ratings should not be seriously taken into account either by A nor by B players, since C players lack motivation for their job. Nevertheless, intrinsic incentives may compensate this lack of material ones.

In the *Fixed incentive* treatment, C players received a fixed payoff of 10 MU, equal to A and B endowments. Fixed incentives mirror different social situations and should motivate C players to reciprocate by increasing their effort in producing careful ratings.

In the *A incentive* treatment, C players' earnings were equal to the payoff of A players. This alignment of interests could resolve the principal-agent problem between intermediaries and trustors, by motivating the agents (intermediaries) to act on behalf of the principals (trustors) guaranteeing that the self-interest of the latter coincides with the objectives of the former. This treatment is therefore expected to lead to more reliable ratings and higher efficiency.

In the *B incentive* treatment, C players' earnings were equal to the payoff of B players. The alignment of trustees' and intermediaries' interests was expected to determine an exploitation of the goodwill of trustors and therefore to produce less reliable ratings and lower trustors' investment.

3.2 Experiment 2 (fixed roles)

A total of 244 subjects (55% females) participated in the second experiment at the Faculty of Economics of the University of Brescia in 2011. Participants were recruited and played as in the first experiment. Overlap with participants in the first experiment has been avoided, so the two experiments could in principle be viewed as a single experiment with a between-subject design. Also the treatments were as before, with the only difference that roles remained fixed throughout the game. Note that the fact that roles no longer alternated actually reduced the sample of observations per role, an effect which is especially relevant in the treatments using the rating system. To overcome this problem, we doubled the number of subjects participating in *No incentive*, *Fixed incentive*, *A incentive* and *B incentive* that were played by 54 subjects each, organized in two sessions encompassing 27 participants.

4 Results

4.1 Experiment 1 (alternating roles)

Previous experiments on the Investment game showed that A players invested on average between one third to half of their endowments. Returns were slightly lower than invest-

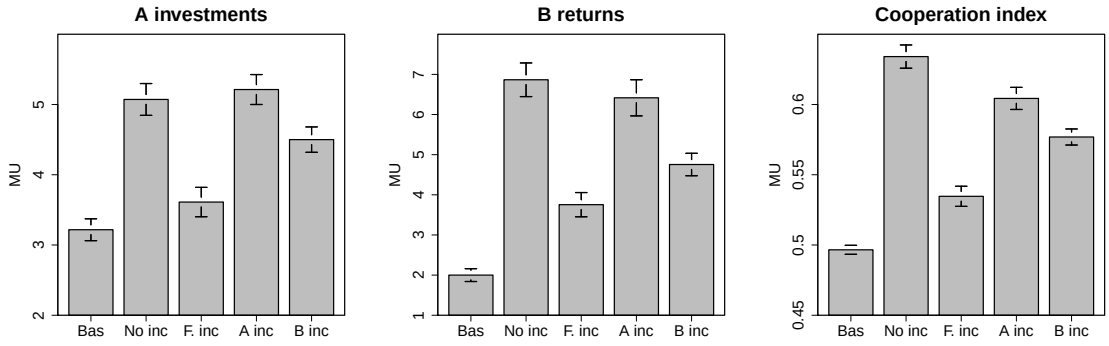

Figure 3: Average investments, returns and CI by treatment in the first experiment with standard error bars.

ments, making trustful behavior not particularly profitable on average (see Johnson and Mislin 2011). Our study replicated these results and, consistently with previous studies which introduced reputational motives in the investment game (Boero et al. 2009a; Keser 2003), showed that the action of intermediaries dramatically improved cooperation. Both investments and returns were higher when intermediaries were introduced, with investments increasing from an average of 3.22 MU in the *Baseline* up to 5.21 MU in *A incentive* and returns rising from 2.00 in the *Baseline* to 6.87 in *No incentive* (Fig. 3). The amounts exchanged in the first three periods of the game, when intermediaries had no previous information to evaluate, and in the last three periods, when B players knew that no further rating would take place, were not included in the analysis.[3]

Differences with the *Baseline* for both investments and returns were significant at the 5% level for all treatments except *Fixed incentive*, where the difference was significant only for returns. Significant differences also existed for B returns. Both *No incentive* and *A incentive* led to higher returns than *Fixed incentive* (Wilcoxon rank sum tests on individual averages, $W = 531.0$, $p = 0.002$ one tailed, and $W = 199.0$, $p = 0.002$ one tailed, respectively). There were no significant differences between *No incentive* and *A incentive* ($W = 385.0$, $p = 0.365$). Differences in investments were smaller, but still remained statistically significant at 5% between *No incentive* and *Fixed incentive* ($W = 508.0$, $p = 0.006$ one tailed) and between *A incentive* and *Fixed incentive* ($W = 176.5$, $p = 0.001$ one tailed).

To describe the dynamics of cooperation in the game better, we built a single indicator that measured cooperation. We did so by looking both at Pareto optimality, which depended only on A investments, and equitable outcome, which depended also on B returns. Specifically, Pareto optimality is an important indicator of the system efficiency in the different treatments. This is indicated by $E = I/d_A$, where I represents A investment and d_A is the endowment. This indicator takes zero when A invested zero and one when A invested the whole endowment. Following previous research (Almås et al. 2010; Fehr and Schmidt 1999; Nowak et al. 2000; Rabin 1993), we considered B returns by following a fairness criterion rewarding outcomes where both players obtained equal payoffs: $F = 1 - [|P_A - P_B|/(P_A + P_B)]$, where P_A and P_B are the payoffs earned by A and B Players

[3]Our dataset may be accessed upon request to the corresponding author. All statistical analyses were performed using the *R* 2.13.0 platform (R Development Core Team 2010).

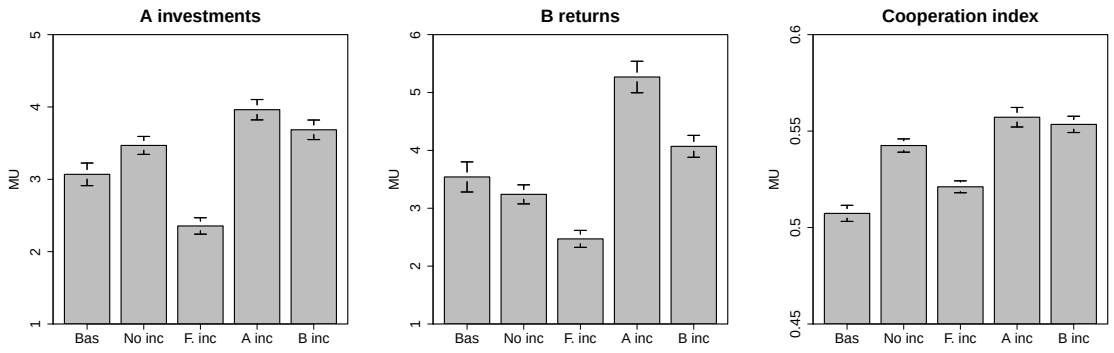

Figure 4: Average investments, returns and CI by treatment in the second experiment with standard error bars.

respectively. This is zero when one of the players obtained the whole amount at stake and the other received zero, while it takes one when both players obtained the same payoff. Averaging the two criteria, we defined the cooperation index as $CI = (E + F)/2$. This takes zero when A players invested zero and B players returned all their endowments, grew coherently with the growth of A investments and a fairer distribution of final payoffs, and reach one when A players invested d_A and B players returned half of their whole endowment, i.e., $(d_B + mI)/2$.

Results showed that the treatment with the highest *CI* was *No incentive*, which led to higher cooperation than any other treatment (Fig. 3). Differences were statistically significant at 10% with *A incentive* and at 5% with all other treatments. The high *CI* value in *No incentive* was especially important as in this case, unlike *A incentive*, intermediaries had no incentive to cooperate with A players.

To sum up, in the first experiment, treatments using intermediaries led to higher trust and trustworthiness than the *Baseline*, with more cooperation in *No incentive* and *A incentive*.

4.2 Experiment 2 (fixed roles)

Generally, the second experiment led to less trust and trustworthiness than the first one. Investments ranged from 2.36 ± 0.11 UM in *Fixed incentive* to 3.96 ± 0.14 UM in *A incentive*. Returns ranged from 2.47 ± 0.15 UM in *Fixed incentive* to 5.27 ± 0.27 UM in *A incentive*. The cooperation index ranged from 0.507 ± 0.004 UM in the *Baseline* to 0.557 ± 0.005 UM in *A incentive* (Fig. 4).

The only treatment leading to investments significantly higher than the baseline was *A incentive* ($W = 82$, $p = 0.049$ one tailed), while returns were significantly higher (at the 10% level) in both *A incentive* and *B incentive* ($W = 88$, $p = 0.071$ one tailed for both treatments). Finally, the cooperation index was significantly higher in *No incentive* ($W = 351$, $p = 0.019$ one tailed), *A incentive* ($W = 299$, $p = 0.003$ one tailed) and *B incentive* ($W = 312$, $p = 0.004$), but not in *Fixed incentive* ($W = 462$, $p = 0.288$ one tailed). To sum up, although professional intermediaries improved cooperation compared to dyadic interaction, *A incentive* was the only treatment outperforming the baseline.

Treatment		Altern. roles		Fixed roles		AR/FR	Wilcoxon	
		Mean	SE	Mean	SE		*W*	*p*
Baseline	A investments	3.22	0.16	3.07	0.16	1.05	207.5	0.385
	B returns	2.00	0.16	3.54	0.26	0.56	142.5	0.079
	B ret. (prop.)	0.09	0.01	0.16	0.01	0.56	137.0	0.059
	CI	0.50	0.00	0.51	0.00	0.98	319.0	0.118
No incentive	A investments	5.07	0.23	3.47	0.12	1.46	174.0	0.029
	B returns	6.87	0.42	3.24	0.16	2.12	196.0	0.003
	B ret. (prop.)	0.24	0.01	0.14	0.01	1.71	189.0	0.006
	CI	0.63	0.01	0.54	0.00	1.17	396.0	0.000
Fixed inc.	A investments	3.61	0.21	2.36	0.11	1.53	170.0	0.040
	B returns	3.75	0.30	2.47	0.15	1.52	142.0	0.232
	B ret. (prop.)	0.17	0.01	0.10	0.01	1.70	150.0	0.153
	CI	0.54	0.01	0.52	0.00	1.04	264.0	0.319
A incentive	A investments	5.21	0.21	3.96	0.14	1.32	174.5	0.028
	B returns	6.42	0.45	5.27	0.27	1.22	147.0	0.180
	B ret. (prop.)	0.23	0.01	0.21	0.01	1.10	134.0	0.333
	CI	0.60	0.01	0.56	0.01	1.07	365.0	0.002
B incentive	A investments	4.50	0.18	3.68	0.14	1.22	154.0	0.121
	B returns	4.75	0.28	4.07	0.19	1.17	141.5	0.238
	B ret. (prop.)	0.19	0.01	0.18	0.01	1.06	130.0	0.387
	CI	0.58	0.01	0.55	0.00	1.05	307.0	0.071

Table 1: Overview of experiment 1 and 2 results (p values are one tailed).

4.3 Comparison of the two experiments

Table 1 shows a comparison between all the alternating vs. fixed roles treatments. Besides lower returns in the baseline—a result consistent with the existing literature (see Johnson and Mislin 2011)—cooperation was generally higher in the alternating role experiment. This difference was especially relevant in *No incentive*. This was one of the best performing treatments in the first experiment, while in the case of fixed role treatments, it only led to cooperation levels similar to the baseline (i.e., without intermediaries). On the other hand, the two treatments with clear monetary incentives led to a more modest decrease in investments and, especially, in returns. Note also that *B incentive* performed similarly in the two experiments.

In order to examine the effect of the different incentive schemes, we estimated a random effects model using dummies indicating each treatment, the fixed role experiment, the first and the last period, and the second half of the game as regressors (Tab. 2). All conditions except, *Fixed incentives* led to higher trust and trustworthiness than the baseline. Fixed roles determined a decrease of trust and cooperation, but not of trustworthiness. Finally, all indicators showed a decline of cooperation during the game, which is consistent with previous studies (e.g. Boero et al. 2009b; Cochard et al. 2004; Keser 2003).

If we consider the behavior of the intermediaries, we can observe that generally they played fairly, asking B players to return significantly more to award a more positive rating (Tab. 3a). They were more demanding in all the treatments where trust was higher, namely

Dependent	Investments	Returns	CI
(Intercept)	3.607^{***}	0.154	0.522^{***}
No incentive	1.339^{**}	1.509^{**}	0.078^{***}
Fixed incentive	0.040	0.684	0.026^{*}
A incentive	1.538^{***}	2.004^{***}	0.075^{***}
B incentive	1.075^{**}	$1.012^{\dagger}$	0.061^{***}
Fixed roles	-0.962^{***}	–0.133	-0.029^{***}
First period	0.066	0.954^{***}	0.025^{***}
Last period	-0.562^{**}	-0.848^{**}	-0.033^{***}
Periods 16–30	-0.313^{***}	-0.420^{***}	-0.009^{***}
A investment		0.853^{***}	
N	4070	4070	8140
F	21.5^{***}	320.7^{***}	61.8^{***}

Table 2: RE regression coefficients. Significance codes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$.

	(a) Return proportion			(b) Rating distribution		
Treatment	Negative	Neutral	Positive	Negative	Neutral	Positive
No incentive AR	0.14	0.22	0.34	0.42	0.22	0.36
No incentive FR	0.13	0.15	0.14	0.37	0.26	0.37
Fixed incentive AR	0.08	0.16	0.24	0.34	0.31	0.35
Fixed incentive FR	0.11	0.09	0.17	0.36	0.30	0.34
A incentive AR	0.15	0.22	0.33	0.41	0.25	0.35
A incentive FR	0.18	0.16	0.28	0.37	0.23	0.40
B incentive AR	0.14	0.20	0.22	0.33	0.29	0.38
B incentive FR	0.19	0.15	0.18	0.33	0.22	0.44
All treatments	0.15	0.16	0.22	0.36	0.26	0.38

Table 3: Average return proportion by rating (a) and rating distribution per treatment (b).

No incentive and *A incentive* in the alternating role experiment. In both cases, to award a positive rating to trustees, intermediaries asked that trustees returned about one third of the trustors' investment. Vice-versa, in treatments less cooperative such as *No incentive* FR or *Fixed incentive* FR, they were less demanding (i.e., to one sixth or even less).

This also applies to rating distribution, even if in this case the differences were less pronounced. Negative ratings were more frequent in the most cooperative treatments. This means that a more rigorous award process took place in these conditions, despite the overall higher trustworthiness of trustees achieved in these treatments (Tab. 3b).

It is also important to note that A player investments reflected the behavior of intermediaries. Generally, subjects invested more when their opponents received positive ratings, less in case of neutral ratings and even less in case of negative ratings (Tab. 4). It is worth noting that also in this case differences across treatments were significant. A players trusted more intermediaries when the latter gave positive ratings in *No incentives* and *A incentive* in the alternating roles experiment. In these cases, they invested on average 6.26 MU and 7.04 MU respectively. They invested significantly less in treatments where there

Treatment	Negative	Neutral	Positive
No incentive AR	3.72	5.76	6.25
No incentive FR	2.39	3.24	4.62
Fixed incentive AR	1.84	4.05	4.81
Fixed incentive FR	2.18	2.44	2.43
A incentive AR	3.43	5.02	7.04
A incentive FR	2.75	4.46	4.71
B incentive AR	3.39	4.52	5.27
B incentive FR	2.79	4.00	4.10
All treatments	2.72	3.87	4.56

Table 4: Average investments by treatment and rating.

was low cooperation, such as *Fixed incentive* in the fixed role condition, where average investment of trustors was as low as 2.43 MU even when trustees had a positive rating.

5 Discussion

Our experiments have highlighted the crucial role of intermediaries for cooperation in trust situations. This can be due to the fact that (i) triadic interactions implied that information available to trustors increased compared to dyadic interactions and that (ii) trustees acted more properly when under observation by a third-party. Intermediaries determined an improvement of the quality of the information that trustors used to discriminate between trustworthy and untrustworthy opponents, as clearly shown in Table 4. This, in turn, provided a rational incentive for trustees to be more trustworthy, as testified by the increase of average B returns in the game.

Indeed, it is expected that when more information about subject behavior circulates in a social system, non-cooperative individuals know that their behavior could be detected and sanctioned by others and so are more tempted to cooperate (Bravo and Tamburino 2008). Recent research showed that this mechanism is relatively independent of the quality of information. For instance, by building an experimentally grounded agent-based model, Boero et al. (2010) showed that the availability of information from other players increased the agents' confidence and their capability of undertaking risky trust investment decisions even if agents had proof that this information was often incorrect.

Returning to our hypotheses, we can say that H1 was confirmed by data. In both experiments, *A incentive* guaranteed high levels of cooperation as intermediaries' evaluations fostered both investments by trustors and returns by trustees. In this treatment, trustees were pushed to pursue reputation building strategies and so returned more. Conversely, trustors thought that intermediaries' opinion was credible and followed them to discriminate between "good" and "bad" opponents. Intermediaries were functional to provide room for this interplay and this pushed the system towards higher cooperation levels.

Contrary to H2, *B incentive* led to higher cooperation than the baseline. To explain this result it is worth noting that, while it is true that intermediaries in *B incentive* were less demanding than in *A incentive* to award positive ratings, they still earnestly discriminated between trustworthy and untrustworthy trustees (Tab. 3). As a consequence, ratings were

useful for trustors who followed them (Tab. 4), even if intermediaries' incentives were misaligned with their interests.

This mismatch between incentives and behavior is extremely relevant. Unfortunately, our experiments cannot help us to look at the reasons behind the intermediaries' behavior and their fairness. In our view, the internalization of social norms about the "fair way of doing things" by intermediaries and their feeling of being under observation by trustors, which are two social mechanisms not mutually exclusive, could account for this. Moreover, the alternating role protocol allowed subjects to follow indirect reciprocity strategies and so they had a serious interest in keeping high the credibility and quality of ratings for future benefits, which also explains most of the differences between the two experiments.

Our argument is supported by previous studies showing that cooperation in typical market situations tends to increase when subjects are (or even only suspect to be) under evaluation by others or are aware that their decisions could be possibly known by others, even if this could not have any material consequences on their payoffs (Bateson et al. 2006; Boero et al. 2009b; Haley and Fessler 2005). This can explain the behavior of trustees, but also why our intermediaries played more or less fairly in all the treatments (Tab. 3). Our explanation is that intermediaries were generally fair, no matter if under material incentives or not, and in case under which incentives, as they knew that their ratings were transmitted to trustors, who not only used them for their investment decisions but also to estimate the intermediaries' reliability. In other words, the triadic interaction typical of our game implied that even intermediaries felt to be under evaluation and consequently were more disposed to fairness.

Consistently with H3, *Fixed incentive* led only to an average level of cooperation. A comparison between *No incentive* and *Fixed incentive* may help to understand this point better. At least in the alternating role experiment, while *No incentive* led to more trust and cooperation, *Fixed incentive* barely improved the baseline. This makes little sense in a rational choice perspective, as in both cases the incentives for intermediaries were not clearly defined. On the other hand, while intrinsic motivations were crucial to motivate intermediaries to formulate reliable ratings in *No incentive*, these aspects were crowded out by material incentives in *Fixed incentive*. This is consistent with the *Motivation crowding theory* (Frey and Jegen 2001) and with a number of studies showing that policies designed for self-interested individuals to increase cooperation in a variety of social and economic situations can actually backfire by undermining individual "moral sentiments" (see Bowles 2008). On the other hand, it is likely that higher earnings could have motivated intermediaries to provide reliable ratings and induced trustors to trust them. The possible explanation is that the magnitude of "fixed incentives" that we provided in the experiment was sufficient to crowd out intrinsic motives of subjects, but not to promote reciprocal and self-interested behavior as happens in typical monetary markets (Ariely et al. 2007). Further work is necessary to examine this hypothesis by comparing treatments where fixed incentives of different magnitudes are offered to intermediaries.

H4 was also confirmed by data, at least in the alternating role experiment. In this case, *No incentive* was in general the best treatment for cooperation. On the other hand, it did not foster trust and, especially, trustworthiness when roles were fixed. This can due to the fact that, by fixing the roles, there was no room for indirect reciprocity strategies for agents. Therefore, we must conclude that intrinsic motivations can keep high the cooperation standard if they are linked, even if only indirectly, to some expected future

benefit. This helps trustors and trustees to believe in the credibility of intermediary ratings. This may also have important implications for the idea of "professionalization" of intermediaries, which is mirrored in our fixed role experiment. In some cases, it could be that professionalization could reduce the credibility of intermediary opinions, even if their ratings were eventually reliable and could be useful for trustors.

Finally, consistently with H5, the comparison of the two experiment testified to the utmost importance of indirect reciprocity motives in trust situations. The idea that indirect reciprocity is fundamental in human societies has been advanced by Nowak and Highfield (2011), who argued that this is one of the most crucial forces in human evolution. Our experiments suggest that, *ceteris paribus*, a significant part of cooperation in trust situations can be explained by indirect reciprocity motives (Tab. 1 and 2). This is consistent with previous studies on the strength of indirect reciprocity based on different interaction schemes, such as the Ultimatum game (Fehr and Fischbacher 2003) and the Indirect reciprocity game (Seinen and Schram 2006).

Obviously, we must conclude that various social mechanisms might affect cooperation in these situations, including social norms, rational incentives, and indirect reciprocity. Theoretically, we could argue that these mechanisms work independently, but there are serious reasons to expect that positive or negative interactions might exist between them in reality. Our experiments provided a good example of this. Indeed, social norms and indirect reciprocity fostered cooperation in *No incentive* with alternating roles, while material incentives crowded out intrinsic motivations in *Fixed incentive*. Moreover, by fixing roles while keeping all equal in the incentive scheme, social norms did not unfold their true power for cooperation, as indirect reciprocity was ruled out. These findings have important implications both for research and policy. First, they give us an idea of the complexity of social systems and the need for further work to disentangle these mechanisms and show how they work in specific circumstances. Secondly, they call for a reconsideration of the traditional incentive-driven policy approach to cooperation as they show that incentive-response behavior of individuals can be an exception while other, even less costly, social mechanisms could ensure good results.

References

Alexander, R. D. (1987). *The Biology of Moral Systems*. Basic Books, New York.

Almås, I., Cappelen, A. W., Sørensen, E. O., and Tungodden, B. (2010). Fairness and the development of inequality acceptance. *Science*, 328:1176–1178.

Ariely, D., Bracha, A., and Meier, S. (2007). Doing good or doing well? image motivations and monetary incentives in behaving pro-socially. *American Economic Review*, 99:545–555.

Bailey, J. P. and Bakos, Y. (1997). An exploratory study of the emerging role of electronic intermediaries. *International Journal of Electronic Commerce*, 1(3):7–20.

Barrera, D. (2005). *Trust in Embedded Settings*. ICS Dissertation Series, Utrecht.

Barrera, D. and Buskens, V. (2009). Third-party effects on trust in an embedded investment game. In Cook, K., Snijders, C., Buskens, V., and Cheshire, C., editors, *eTrust: Forming Relationships in the Online World*, pages 37–72. Russell Sage, New York.

Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2:412–414.

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

Boero, R., Bravo, G., Castellani, M., Laganà, F., and Squazzoni, F. (2009a). Pillars of trust: An experimental study on reputation and its effects. *Sociological Research Online*, 14(5):5.

Boero, R., Bravo, G., Castellani, M., and Squazzoni, F. (2009b). Reputational cues in repeated trust games. *Journal of Socio-Economics*, 38(6):871–877.

Boero, R., Bravo, G., Castellani, M., and Squazzoni, F. (2010). Why bother with what others tell you? an experimental data-driven agent-based model. *Journal of Artificial Societies and Social Simulation*, 13(2):6.

Bohnet, I. and Frey, B. S. (1999). The sound of silence in prisoner's dilemma and dictator games. *Journal of Economic Behavior and Organization*, 38:43–57.

Bowles, S. (2008). Policies designed for self-interested citizens may undermine "the moral sentiments": Evidence from economic experiments. *Science*, 320:1605–1609.

Bravo, G. and Tamburino, L. (2008). The evolution of trust in non-simultaneous exchange situations. *Rationality and Society*, 20(1):85–113.

Burt, R. S. and Knez, M. (1995). Kinds of third-party effects on trust. *Rationality and Society*, 7(3):255–292.

Buskens, V. and Raub, W. (2008). Rational choice research on social dilemmas: embeddedness effects on trust. In Wittek, R., Snijders, T., and Nee, V., editors, *Handbook of Rational Choice Social Research*. Russell Sage, New York.

Buskens, V., Raub, W., and van der Veer, J. (2010). Trust in triads: An experimental study. *Social Networks*, 32:301–312.

Camerer, C. F. and Weigelt, K. (1988). Experimental test of a sequential equilibrium reputation model. *Econometrica*, 56:1–36.

Cason, T. N. and Mui, V. (1997). A laboratory study of group polarization in the team dictator game. *Economic Journal*, 107:1465–1483.

Cochard, F., Nguyen Van, P., and Willinger, M. (2004). Trusting behavior in a repeated investment game. *Journal of Economic Behavior and Organization*, 55:31–44.

Coleman, J. S. (1990). *Foundations of Social Theory*. Harvard University Press, Harvard.

Cook, K., Hardin, R., and Levi, M. (2005). *Cooperation Without Trust*. Russell Sage Publications, New York.

Cook, K. S. and Hardin, R. (2001). Norms of cooperativeness and networks of trust. In Hechter, M. and Opp, K.-D., editors, *Social Norms*, pages 327–347. Russel Sage Foundation, New York.

Dasgupta, P. (1988). Trust as a commodity. In Gambetta, D., editor, *Trust, Making and Breaking Cooperation*, pages 49–72. Blackwell, Oxford.

Fehr, E. and Falk, A. (2002). Psychological foundations of incentives. *European Economic Review*, 46:687–724.

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature*, 525:785–791.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114:817–868.

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

Frey, B. and Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, 15:589–611.

Gneezy, U. and Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies*, 29(1):1–18.

Greiner, B. (2004). An online recruitment system for economic experiments. In Kremer, K. and Macho, V., editors, *Forschung und Wissenschaftliches Rechnen 2003*, pages 79–93. Ges. für Wiss. Datenverarbeitung, Göttingen.

Hales, D. (2000). Cooperation without space or memory: Tags, groups and the prisoner's dilemma. In Moss, S. and Davidsson, P., editors, *Multi-Agent Based Simulation*, pages 157–166. Springer Verlag, Berlin-Heidelberg.

Haley, K. J. and Fessler, D. M. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26:245–256.

Hardin, R. (2004). *Trust and Trustworthiness*. Russell Sage Publications, New York.

Heyman, J. and Ariely, D. (2004). Effort for payment: A tale of two markets. *Psychological Science*, 15(11):787–793.

Holmaas, T. H., Kjerstad, E., Luraas, H., and Straume, O. R. (2010). Does monetary punishment crowd out pro-social motivations? *Journal of Economic Behavior and Organization*, 75(2):261–267.

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

Keser, C. (2003). Experimental games for the design of reputation management systems. *IBM Systems Journal*, 42:498–506.

Kim, J.-W. (2010). A tag-based evolutionary prisoner's dilemma game on networks with different topologies. *Journal of Artificial Societies and Social Simulation*, 13(3):2.

Laffont, J.-J. and Martimort, D. (2002). *The Theory of Incentives: The Principal-Agent Model*. Pri, Princeton.

Malinowski, B. (1922). *Argonauts of the Western Pacific*. Routledge, London.

Mauss, M. (1954). *The Gift: Forms and Functions of Exchange in Archaic Societies*. Cohen & West, London.

McCabe, K. A. and Smith, V. L. (2000). A comparison of naïve and sophisticated subject behavior with game theoretic predictions. *Proceedings of the National Academy of Sciences USA*, 97:3777–3781.

Meier, S. (2007). Do subsidies increase charitable giving in the long run? matching donations in a field experiment. *Journal of European Economic Association*, 5:1203–1222.

Mellström, C. and Johannesson, M. (2008). Crowding out in blood donation: Was Titmuss right? *Journal of European Economic Association*, 6:845–863.

Milinski, M., Semmann, D., and Krambeck, H.-J. (2002). Reputation helps solve the tragedy of the commons. *Nature*, 415:424–426.

Nowak, M. A. and Highfield, R. (2011). *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed*. Free Press, New York.

Nowak, M. A., Page, K. M., and Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science*, 289:1773–1775.

Nowak, M. A. and Sigmund, K. (1998a). The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, 194:561–574.

Nowak, M. A. and Sigmund, K. (1998b). Evolution of indirect reciprocity by image scoring. *Nature*, 393:573–577.

Palmer, J. W., Bailey, J. P., and Faraj, S. (2000). The role of intermediaries in the development of trust on the WWW: The use and prominence of trusted third parties and privacy statements. *Journal of Computer Mediated Communication*, 5:3.

Polanyi, K. (1968). *Primitive, archaic and modern economies: Essays of Karl Polanyi*. Anchor Books, Garden City, NY.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 83:1281–1302.

Resnick, P. and Zeckhauser, R. (2002). Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. *The Economics of Internet and E-Commerce*, 11:127–157.

Rigdon, M., Ishii, K., Watabe, M., and Kitayama, S. (2009). Minimal social cues in the dictator game. *Journal of Economic Psychology*, 30:358–367.

Seinen, I. and Schram, A. (2006). Status and group norms: Indirect reciprocity in a helping experiment. *European Economic Review*, 50:581–602.

Snijders, C. and Kerens, G. (1998). Determinants of trust. In Budescu, D. V., Erev, I., and Zwick, R., editors, *Games and Human Behavior: Essays in Honor of Ammon Rapoport*, pages 355–385. Lawrence Erlbaum, Mahwah, NJ.

Sommerfeld, R. D., Krambeck, H.-J., and Milinski, M. (2008). Multiple gossip statements and their effect on reputation and trustworthiness. *Proceedings of the Royal Society B*, 275:2529–2536.

Sommerfeld, R. D., Krambeck, H.-J., Semmann, D., and Milinski, M. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Science USA*, 104(44):17435–17440.

Sorenson, O. and Stuart, T. E. S. (2001). Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology*, 106(6):1546–1588.

Swedberg, R. (2005). Conflicts of interests in the US brokerage industry. In Knorr Cetina, K. and Preda, A., editors, *The Sociology of Financial Markets*, pages 187–203. Oxford University Press, Oxford.

Takahashi, N. (2000). The emergence of generalized exchange. *The American Journal of Sociology*, 105(4):1105–1134.

Titmuss, R. M. (1971). *The Gift Relationship: From Human Blood to Social Policy*. Pantheon, New York.