Whatever you say, your reputation precedes you: observation and cheap talk in the trust game

Juergen Bracht

University of Aberdeen Business School
Edward Wright Building
Aberdeen AB24 3QY, UK

Nick Feltovich*

University of Aberdeen Business School
Edward Wright Building
Aberdeen AB24 3QY, UK

Abstract

Behavior in trust games has been linked to general notions of trust and trustworthiness, important components of social capital. In the equilibrium of a trust game, the investor does not invest, foreseeing that the allocator would keep all of the returns. We use a human–subjects experiment to test the effects of changes to the game designed to increase cooperation and efficiency. We add a pre–play stage in which the investor receives a cheap–talk message from the allocator, observes the allocator's previous decision, or both. None of these changes alter the game's theoretical predictions. We find that allowing observation results in substantially higher cooperation and efficiency, but cheap talk has little effect.

Key words: experiment, trust game, cheap talk, observation, mechanism.

*Corresponding author. Financial support from the University of Aberdeen, the University of Houston, and the British Academy is gratefully acknowledged. We thank Todd Kaplan, Andreas Ortmann, Daniel Zizzo, an anonymous referee, and participants at several conferences and seminars for helpful suggestions and comments.

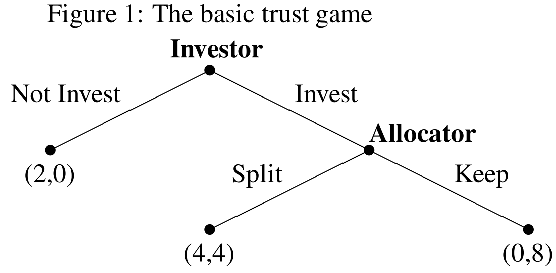
Email addresses: juergen.bracht@abdn.ac.uk (Juergen Bracht), n.feltovich@abdn.ac.uk (Nick Feltovich)

URL: www.abdn.ac.uk/~pec214/ (Nick Feltovich)

1. Introduction

Many researchers have studied the role of "social capital" in societies. Multiple definitions of social capital have been proposed, but a typical one is provided by Putnam (1999): "features of social life, networks, norms, trust that enable participants to act together more effectively to pursue shared objectives". Social capital has been linked with, inter alia, educational outcomes (Coleman (1988)), economic success (Fukuyama (1995), Knack and Keefer (1997)), the strength of political and judicial institutions (Putnam (1993), La Porta et al. (1997)), and the effectiveness and sustainability of development projects (World Bank (1999)). Nearly all definitions of social capital posit that two of its primary components are *trust* (the belief that others act in the interest of some measure of fairness or social welfare rather than their own self–interest) and *trustworthiness* (the extent to which trust in a person is warranted). While measuring attributes such as trust and trustworthiness is inherently open to some subjectivity, recent work has utilized individuals' behavior in simple games to construct such measures. One of the most widely–used games for this purpose is the *trust game* (Berg et al. (1995)), played by two players, who we will call the *investor* and the *allocator*. In a simple version (see Figure 1), the investor chooses whether to pass a sum of money ("Invest") to the allocator. If (and only if) she does so, this sum multiplies in value (for example, as a successful investment) but the allocator has complete discretion over the proceeds: he can either keep everything for himself or return half to the investor. Under the usual assumptions used by economic theorists, the equilib-

Figure 1: The basic trust game



rium prediction for this game is simple.[1] The allocator would always prefer to keep all proceeds from investment rather than returning anything to the investor; understanding this, the investor will invest nothing.

The trust game is an example of a *social dilemma*: individual self–interested behavior leads to an outcome that is inefficient from the perspective of the group. A choice by the investor to pass money to the allocator is not rational from the standpoint of game theory, but rather reflects *trust* that this choice—which benefits the allocator irrespective of his subsequent decision—will in turn be rewarded by the allocator's splitting the proceeds. Accordingly, a choice by the allocator to split the proceeds reflects *trustworthiness*, not self–interest. In fact, previous research has found that trust and trustworthiness measured according to observed behavior in this game are positively associated with trust and trustworthiness measured by responses to attitudinal survey questions. Glaeser et al. (2000) found a positive correlation between the amount allocators return in a trust game and their agreement with statements such as "generally speaking, most people can be trusted" (though they found only a weak relationship between investor behavior and survey responses). Even more significant, Glaeser et al. found positive correlations between both investor and allocator choices and (self–reported) past trusting behavior, measured by responses to questions like "How often do you lend money to your friends?" (p. 819). These correlations suggest that investment by investors and returns by allocators in trust games have some external validity as general measures of trust and trustworthiness.

Because of the close connection between behavior in trust games and desirable social outcomes, there has been interest in how such games (and social dilemmas in general) can be modified in order to improve their outcomes. Some researchers

[1]Throughout this paper, we will use terminology such as "equilibrium prediction" to mean the combination of appropriate equilibrium concepts and the assumption that preferences concern only players' own monetary payoffs. We acknowledge that this is an abuse of terminology, as game theory itself makes no assumptions about what form preferences take, and if preferences involve non–monetary aspects, the true equilibrium predictions may be different. In Section 4.5, we discuss the implications of alternative assumptions about preferences or cognitive ability.

have examined mechanisms that change the theoretical predictions of the game, so that efficient outcomes become equilibrium outcomes (see Bracht and Feltovich (2008) for a survey). Experimental tests generally find that such mechanisms do improve outcomes (though often less than predicted due to "crowding out").[2] However, these mechanisms can be unsatisfying from the standpoint of application, since most rely on a third party with coercive power, either to impose changes to the game's structure or payoffs (e.g., Falkinger et al. (2000), Van Huyck et al. (1995)), or to enforce an agreement made by the players (Andreoni and Varian (1999), Bracht and Feltovich (2008)). The reason this is unsatisfying is that the situations where trust games are likely to arise are typically exactly those where such a third party either does not exist or has circumscribed powers. (Otherwise, why not have it simply impose the preferred outcome in the basic game, without introducing the complication of the mechanism?)

Therefore, a natural next step is to consider "non–coercive" mechanisms: modifications that do not require a third party with coercive power. We focus on one class, which we call "information mechanisms"; these involve no changes to the game other than opportunities to give or receive information. In our setup, this information can take one of two forms: cheap talk from the allocator to the investor, and observation of the allocator's previous action. These two information mechanisms have several desirable characteristics. First, though they may not eliminate the requirement of a third party (for example, accurate observation would seem to rely on one), any third party need not have coercive power. Second, they do not require a central authority with precise information about players' preferences, nor do they require the players themselves to have such information. Finally, use of the mechanism does not cause efficiency losses (in contrast to costly punishment

[2] A common result in mechanism experiments is that observed levels of cooperative behavior are higher than predicted when no mechanism is in place, but at or even below the prediction when there is a mechanism. Ostrom (2000), summarizing a large body of relevant research, concluded that "externally imposed rules tend to 'crowd out' endogenous cooperative behavior" (p. 147). See also Lazzarini et al. (2004), who discuss recent research on crowding out, including experiments in which crowding out was not seen.

mechanisms).

In order to assess the performance of these information mechanisms, we designed and ran a human–subjects experiment with several treatments. Each treatment comprises five rounds of a standard trust game, followed by ten more rounds. In our baseline treatment, these next ten rounds are also of the trust game. In the other treatments, these rounds use a modified trust game with cheap talk but no observation, one with observation but no cheap talk, or one with both. Standard game theory predicts that none of these treatments should alter either player's behavior. Contrary to this prediction, our main result is that introducing observation—of even just one previous action—improves aggregate cooperation and efficiency substantially, and significantly, over the basic trust game (in contrast to other studies where observation involves large parts of the opponent's history of play). At the individual level, allocators' previous actions are positively correlated with their current–round actions, and possibly as a result, investors are more likely to invest when faced with an allocator who was observed to have behaved cooperatively. These effects disappear in the final round, when subjects know that actions will not be observed in the future. By contrast, cheap talk has little effect on behavior by either player in any round.

2. Theory and implications

The basic game is the one shown in Figure 1. In the unique subgame perfect equilibrium, the investor does not invest, correctly anticipating that nothing would be returned if she did. Experimental trust–game studies, on the other hand, generally find these main results: investors often invest; allocators return a positive amount with nonnegligible frequency; amounts returned average about the same as, or somewhat below, amounts invested (so that investors typically do not earn a profit by investing); and when play of the game is repeated, amounts invested and returned trend toward zero. These results are quite robust to variations in the experimental design (see Bracht and Feltovich (2008) for a discussion).

Though observed levels of trust and trustworthiness are typically well above zero, they are also well below 100%, so it is of interest to examine modifications that may improve outcomes. One is allowing *cheap talk*: a costless nonbinding message from the allocator to the investor prior to the investor's decision, indicating the action the allocator intends to take should the investor choose to invest. The other modification we consider is *observation* of the allocator's previous action—whether the allocator chose Keep or Split the most recent time he was faced with that decision. In our experiment, players are rematched after every round, meaning the previous action will have been chosen when the allocator was matched with a *different* investor, so that the information gained by observation is genuinely new (that is, it does not simply restate information the investor already had).

Our experiment involves finite repetition of four versions of the trust game. One is the basic game. A second is our "W" (words) game, with cheap talk but no observation. In our "D" (deeds) game, there is observation but no cheap talk, and our "WD" (words and deeds) game has both cheap talk and observation. As noted in the introduction, the theoretical prediction for all of these variations is the same as that for the basic game.

Many researchers have looked experimentally at the effects of either cheap talk or observation in social dilemmas. The effect of cheap talk seems to depend greatly on what form it takes. When subjects can communicate face–to–face, with few or no limits on content, cooperation and efficiency can be substantially enhanced (Isaac and Walker (1988)). On the other hand, when communication takes place through a computer system and is limited to single letters or numbers, it typically leads to only a minor improvement (Duffy and Feltovich (2002)), no systematic effect (Bochet et al. (2006)), or even a deleterious one (Wilson and Sell (1997)). Along these lines, Sally (1995)'s meta–analysis of prisoners'–dilemma experiments finds significant positive relationships between the frequency of cooperation and both the fraction of rounds before which verbal communication was allowed and the elicitation of promises from subjects, but not between cooperation and written communication. Also, Bochet et al. (2006) found higher public good contributions under a "chat room" (no restriction on messages) treatment than when only numerical messages were possible, with still higher contributions in a face–to–face treatment.

The literature on the effects of observation suffers somewhat from a lack of agreement on what constitutes observation—or more precisely, the fact that observation can serve several purposes. One strand looks at the effects of giving players information about others in the same role, with the idea that successful behavior might be imitated (Duffy and Feltovich (1999), Huck et al. (2000)). A second treats observation similarly to end–of–round feedback, which might actually reduce cooperation (Wilson and Sell (1997)). A third considers observation by a third party, often with the ability to reward or punish uncooperative players (Kahneman et al. (1986), Fehr and Fischbacher (2004)). Our treatment of observation is yet another kind: information about a current opponent's past behavior, with no opportunity for reward or punishment except via the game itself. Duffy and Feltovich (2002) found that this kind of observation led to increased cooperation and efficiency in a prisoners' dilemma. Keser (2004) considered rating systems closely related to observation. Following play of a trust game, the investor rated the allocator as positive, neutral, or negative. In a "short–run reputation" treatment, prior to play, the investor would observe the rating from the most recent round; in a "long–run reputation" treatment, the entire history of ratings would be observed. Keser found that adding either mechanism led to increases in cooperative behavior, and her results suggest that the long–run mechanism may increase cooperation more than the short–run mechanism.[3]

We know of only one paper attempting a direct comparison of cheap talk and the kind of observation we study: Duffy and Feltovich (2006), who did so using a two–player prisoners' dilemma. They found that allowing either observation or cheap talk resulted in higher levels of cooperation than when neither was available, but that the incremental social benefit of the second type of information was negligible. However, their use of a simultaneous–move game made drawing in-

[3]Bolton et al. (2004) and Huck et al. (2006) allow investors to see allocators' entire histories under a random–matching protocol. Huck et al. find that cooperation is no more likely with this extra information than without it. Bolton et al. do find that cooperation increases, though less so than under fixed–pairs matching.

ferences from observed actions difficult. For example, a player observing that her opponent defected in the previous round could not distinguish among several reasons for this choice—opportunistic behavior, punishing his previous opponent for an earlier defection, or perhaps a belief that the previous opponent was also going to defect. In our setup, by contrast, only allocators' previous actions are observed, so they have no scope to punish opponents for past actions. Moreover, allocators face no strategic uncertainty, so there is never a question of what they believe their current opponent will do. Thus, there is much less room for investors in our experiment to misconstrue an opponent's previous action.

3. Experimental design and procedures

Each session comprised 15 rounds, beginning with five rounds of the basic trust game, intended to familiarize subjects with the strategic situation and the computer interface. The remaining ten rounds depended on the treatment. In the C (control) treatment, subjects continued playing the basic trust game. In the W, D, and WD treatments, they played the W, D, and WD games respectively (see Table 1). Sessions typically involved 20 subjects, though some had fewer due to no–shows. Subjects were primarily undergraduate students from University College London and Exeter University, and were recruited by a variety of methods. No one took part in more than one session.

Table 1: Treatments used in the experiment

Treatment	Rounds	Cheap talk?	Observation of previous action?	Number of sessions	Total number of subjects
All	1–5	No	No	13	240
C	6–15	No	No	3	60
W	6–15	Yes	No	4	78
D	6–15	No	Yes	3	54
WD	6–15	Yes	Yes	3	48

At the beginning of a session, subjects were seated in a single room and given written instructions for the first five rounds.[4] No specifics about the later rounds were given at this point, other than that the second part of the session might be the same or different. The instructions were read aloud by the experimenter, after which the first round of play begun. After the fifth round was completed, subjects were given instructions for the remaining ten rounds; these were read aloud, then those rounds were played. Importantly, the instructions stated both the number of rounds that would be played—so that reading them aloud should bring about common knowledge of the (finite) endpoint of the game—and the nature of observation, so that allocators always knew exactly when their actions would be observed in the next round, and investors were aware of this as well.

The experiment was run on networked computers using z–Tree (Fischbacher (2007)). Subjects were randomly assigned to roles at the beginning of a session and kept these roles for all rounds. In rounds 1–5, each investor was matched to each allocator at most once. In sessions with 20 subjects, each investor was matched exactly once in rounds 6–15 to each allocator; in sessions with fewer subjects, investors and allocators could be matched more than once.[5]

The sequence of play in a round was as follows. If cheap talk was permitted, a round began with the allocator being prompted to choose a message, which would then be seen by the investor.[6] At that time, the investor also observed the allocator's previous action, if that information was available.[7] After cheap

[4]The instructions used in the experiment, as well as the raw data, are available from the corresponding author upon request.

[5]We tried to reduce the possibility of supergame effects by not telling subjects the ID number of their counterparts, so that they would not be certain sure when they were facing someone whom they had faced previously.

[6]One design issue is whether allocators should be allowed to send no message at all. In half of our cheap–talk sessions, we did allow "blank messages", while in the other half, we required messages to be either Keep or Split. In the data, we found no systematic differences between the two types of cheap–talk treatment, so we ended up pooling all of the cheap–talk sessions.

[7]Because the allocator makes a Keep/Split choice only when his counterpart chooses Invest, it might be that the previous action was not in the previous round. As long as the allocator had made at least one Keep/Split decision in or after round 6, the investor was told the allocator's most recent

talk and observation had taken place, investors were prompted to choose Invest or Not Invest, after which each allocator would see his counterpart's decision and (if it was Invest) be prompted to choose Split or Keep. Then, all subjects received end–of–round feedback: any pre–play information, the investor's choice, the allocator's choice (if applicable), and the subject's own payoff. Subjects were not explicitly told their counterparts' payoffs, though they could calculate them easily if they wished. Subjects were asked to observe their results and write them into a record sheet, then click a button to continue to the next round.

At the end of the session, subjects were paid their earnings in cash. Earnings consisted of a £5 show–up fee (at the time, worth roughly $9), plus payoffs from two randomly chosen rounds (one from the first five and one from the last ten) at an exchange rate of £1 per point. Average earnings were roughly £10 for a session typically lasting about 45 minutes.

4. Experimental results

The experiment consisted of thirteen sessions: four of the W treatment and three of each other treatment. We begin our discussion of the results by presenting summary statistics characterizing aggregate behavior in each of our treatments. Later, we will look at round–by–round behavior and examine estimation results based on parametric statistical models.

4.1. Aggregate choice frequencies

Table 2 shows that in rounds 1–5, when everyone played the basic trust game, frequencies of Invest and Split are similar across cells.[8] Overall frequencies of

---

choice, but not the round in which it took place. It was also possible that there was no previous action to observe, if the allocator had no such decisions since round 6. In this case, the investor was informed that the allocator had no previous actions to observe. This was fairly uncommon; 12 out of 51 allocators in round 7 of the D and WD treatments had no such previous action, 4 in round 8 had none, and from round 9 on, all had a previous action.

[8] Chi–square tests fail to reject the null hypothesis that frequencies of Invest in rounds 1–5 are equal across the four cells ($\chi^2 \approx 2.19$, d.f.=3, $p \approx 0.53$), and likewise for frequencies of

Invest choices (0.552) and Split choices (0.438) in these rounds are comparable to those from other trust–game studies, though a bit on the high side—perhaps because the (Invest, Split) outcome is both efficient and fair.[9]

Table 2: Aggregate subject behavior

Sessions	Frequency of Invest				Conditional frequency of Split			
	Rounds 1–5		Rounds 6–15		Rounds 1–5		Rounds 6–15	
C	0.567	(85/150)	0.400	(120/300)	0.376	(32/85)	0.408	(49/120)
W	0.585	(114/195)	0.405	(158/390)	0.482	(55/114)	0.399	(63/158)
D	0.519	(70/135)	0.633	(171/270)	0.386	(27/70)	0.825	(141/171)
WD	0.517	(62/120)	0.712	(171/240)	0.500	(31/62)	0.877	(150/171)

Behavior in rounds 6–15, on the other hand, shows substantial variation across treatments. For both roles, the cooperative action (Invest or Split) is much more likely in treatments with observation (D and WD) than without (C and W). Robust rank–order tests on session–level data show these differences to be significant (pooled D and WD versus pooled C and W, $p < 0.01$ for investors, $p < 0.001$ for allocators). In contrast, cheap talk has relatively little effect on the likelihood of the cooperative action, whether or not observation is possible; even pooling the C and D treatments and the W and WD treatments, we find no significant differences for either role ($p > 0.10$). If, instead of comparing levels of cooperative behavior in rounds 6–15, we compare the differences between these levels and their corresponding levels in rounds 1–5 (in an effort to control for intrinsic differences in cooperativeness across subjects), we get similar results: a substantial effect from

Split ($\chi^2 \approx 3.97$, d.f.=3, $p \approx 0.26$). See Siegel and Castellan (1988) for descriptions of the nonparametric tests used in this paper, and see Feltovich (2005) for the critical values used later for the robust rank–order test. We note that failure to reject a null hypothesis of equality is typically weak evidence that the frequencies are actually equal; however, not only are these p–values far from conventional levels of significance, but the test we've used is extremely liberal: it ignores dependence among subjects in the same session and for a given subject across rounds.

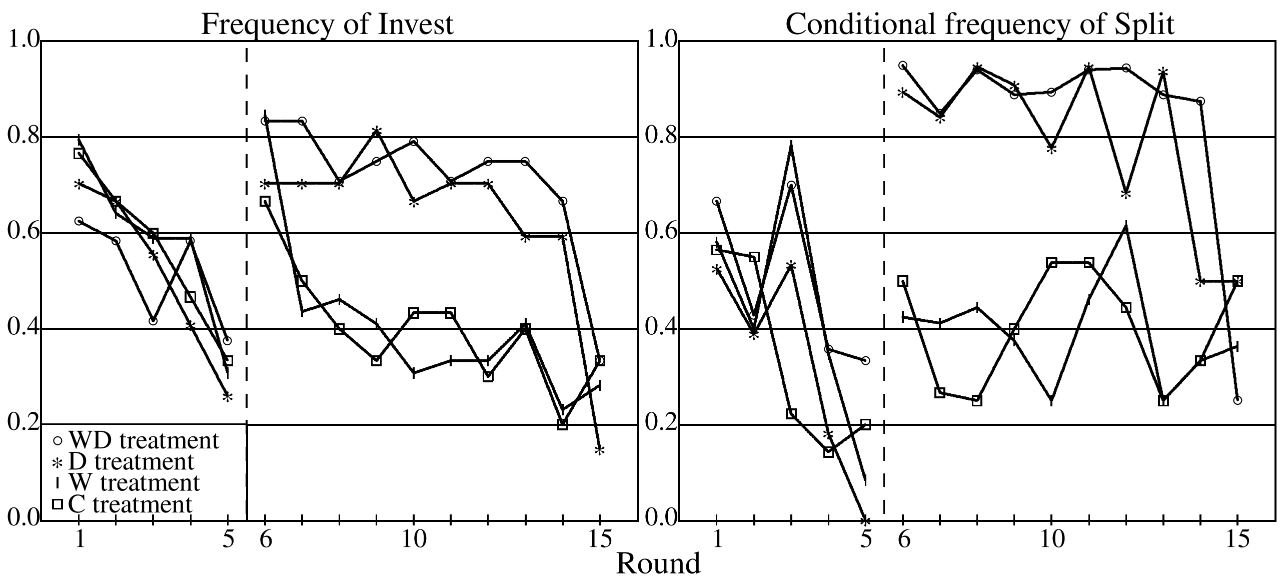
[9]We thank an anonymous referee for pointing this out.

observation ($p \approx 0.001$ for investors and $p < 0.001$ for allocators for pooled C and W versus pooled D and WD), but a negligible effect from cheap talk ($p > 0.10$ for both roles for pooled C and D versus pooled W and WD).

4.2. *Round–by–round behavior*

Figure 2 shows that in rounds 1–5, the frequencies of Invest and Split begin at a high level, but drop sharply over time. We see a restart effect (Andreoni

Figure 2: Subject behavior in each round



(1988)) in the C treatment; frequencies of Invest and Split jump sharply upward from round 5 to round 6, even though no feature of the game has changed.[10] The corresponding frequencies also jump upward in the other treatments, though it is unclear in these cases whether this is a restart effect or the result of changes to the game's structure and (perceived) incentives.

[10]For other examples of restart effects, see Camerer and Fehr (2003) and Duffy and Ochs (2009). In particular, Duffy and Ochs found a restart effect even under random matching, as we do.

In the C and W treatments, the frequency of Split is about one–half after the restart and remains constant over time; in the D and WD treatments, it is substantially higher (between 80% and 90%) for several rounds, until dropping precipitously over the last one or two rounds to about the level seen in the C and W treatments.[11] The frequency of Invest is roughly similar across treatments in round 6 (despite the differences seen in Split frequencies), but diverges quickly. In the C and W treatments, this frequency drops quickly at first, then gradually. In the D and WD treatments, on the other hand, the frequency of Invest stays roughly constant for several rounds before dropping sharply at the end of the session.

The patterns of behavior in rounds 6–15 can thus be summarized as follows. Frequencies of Invest and Split are usually substantially above zero (the theoretical prediction). When observation is not possible, the frequency of Split is relatively low, usually below the level of one–half that is needed to make Invest a monetary best response for investors; perhaps in response, the frequency of Invest declines steadily over time. On the other hand, when observation is possible, frequencies of Invest and Split start high and stay high until shortly before the end of the game, at which point both drop sharply.

We next use probit regressions to look more closely at the effects of cheap talk and observation in rounds 6–15. For our first specification, the dependent variable is an indicator for Invest (1 = Invest, 0 = Not Invest). In order to control for various aspects of time dependence, we include a right–hand–side variable for the round number (normalized to range from 1 to 10), an indicator whose value is 1 in the last round of all treatments, and one whose value is 1 in the last round of a treatment with observation. To examine the effects of our treatments, we include indicators for cheap talk available, observation available, and both available, as well as products of each of these with the (normalized) round number. Finally, as an attempt to control for individual differences in intrinsic trust, we include a measure of trust from the first part of the session; for investors, this is just the

[11] Sharp dropoffs in cooperative behavior in a publicly–announced last round of a repeated social dilemma are common. See, for example, Fehr et al. (1997) and Riedl and Tyran (2005). See also our discussion in Section 4.5.

frequency of Invest choices in rounds 1–5. For our second specification, the dependent variable is an indicator for Split, and we restrict our data to the subset following an Invest choice. We use the same right–hand–side variables as in the first specification, except that the trust measure is replaced by a measure of trustworthiness: the frequency of Split choices in rounds 1–5.[12] Both specifications were estimated using Stata (version 10) with individual–subject random effects.

Table 3 shows coefficient estimates and standard errors for each variable in both models, as well as log–likelihoods and pseudo–R^2s for both specifications.[13] The results confirm what we saw in the descriptive statistics. For investors, the negative coefficient for the round number indicates that cooperative behavior declines over time (though the coefficient for allocators, while also negative, is not significantly different from zero). The negative and significant value for the "D*[final round]" indicator for both roles captures the drop in levels of Invest and Split in the final round of the games with observation—when subjects know that the allocator's action in the current round will not be observed in any future round.

4.3. Comparison of observation and cheap talk

Table 3 suggests that observation has a large effect on behavior, while the effect of cheap talk is at best substantially smaller, and possibly negligible. We now estimate expressions for the incremental effects of allowing observation or cheap talk on the probability of choosing Invest or Split, using the model specifications from Table 3. As an example, if β_{W} and $\beta_{\mathrm{W*round}}$ are the coefficients of W and W*round respectively, then the incremental effect of the W game instead of the basic game on the probability of Invest or Split in round t is $\Phi\left(\bar{X} \cdot B + \beta_{\mathrm{W}} + \beta_{\mathrm{W*round}} \cdot t\right) - \Phi\left(\bar{X} \cdot B\right)$, where Φ is the standard normal cumulative distribution function, $\bar{X}$ is the row vector of the other right–hand variables' values (in the following, either unconditional sample means or means

[12] One of the allocators in the experiment faced no Invest choices in rounds 1–5, so this measure is undefined; for this allocator, we used the aggregate Split frequency from rounds 1–5.

[13] Pseudo–R^2 values were computed by rescaling the log–likelihoods into [0,1], with a model with only the constant term on the right–hand side mapping to zero, and a perfect fit to one.

Table 3: Coefficients from probit models with random effects (standard errors in parentheses)

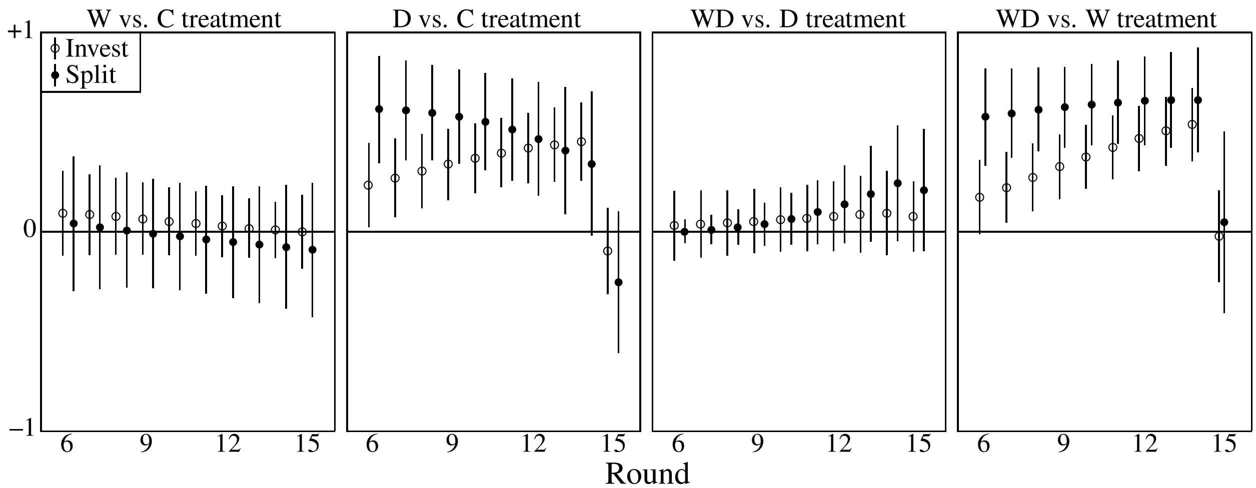
Dependent variable	Invest ($N = 1500$)	Split (given Invest) ($N = 854$)
constant	–0.034	–0.327
	(0.338)	(0.401)
round number	-0.136^{***}	–0.020
	(0.033)	(0.054)
final round	0.381^{*}	0.040
	(0.223)	(0.434)
measure of trust/	0.572	–0.005
trustworthiness	(0.374)	(0.402)
W (any game with cheap talk)	0.273	0.148
	(0.309)	(0.491)
W*round	–0.027	–0.043
	(0.040)	(0.067)
D (any game with observed actions)	0.611^{*}	2.561^{***}
	(0.347)	(0.602)
D*round	0.071	-0.186^{**}
	(0.047)	(0.081)
D*[final round]	-1.672^{***}	-1.774^{**}
	(0.341)	(0.699)
WD (cheap talk + observed actions)	–0.173	–0.230
	(0.486)	(0.808)
WD*round	0.045	0.146
	(0.061)	(0.108)
–ln(L)	670.447	268.399
pseudo–R^2	0.093	0.149

* (**,***): *Coefficient significantly different from zero at the 10% (5%, 1%) level.*

conditional on an appropriate sub–sample), and B is the column vector of their coefficients.

Figure 3 shows point estimates and 95% confidence intervals for the incremental effects for the W and D games versus the basic trust game, and the WD game versus the W and D games. This figure highlights the sharp difference in efficacy between observation and cheap talk. The first and third panels show that

Figure 3: Estimated incremental effects of treatment on Invest/Split choice *(Circles represent point estimates; line segments represent 95% confidence intervals)*

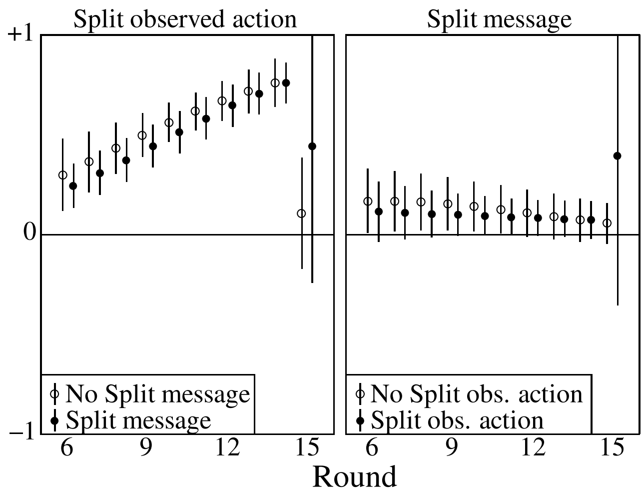


adding cheap talk has no significant effect on cooperation by either player, irrespective of whether observation is possible. On the other hand, the second and fourth panels show that adding observation—either when cheap talk is available or when it is not—nearly always results in significantly more cooperation by both types of player. The exception is round 15, when there is no significant effect.

Disaggregating the data further tells a largely similar story. The left panel of Figure 4 shows that observing a Split action tends to increase the likelihood of the investor's choosing Invest in the current round.[14] The effect grows over time until plummeting in the final round—both when the investor had also received a Split message and when she did not—and is significant in all but this final round. The right panel shows that the point estimate of the effect of receiving a Split cheap

[14]The incremental effects in this figure are based on estimation of a new random–effects probit model, whose dependent variable is an indicator for Invest. Right–hand–side variables are the round number, the measure of trust used in Table 3, indicators for Split message, Split observed action, and both Split message and Split observed action, and products of these three indicators with the round number and with a final–round indicator.

Figure 4: Estimated incremental effects of messages and observed actions on Invest choices *(Circles represent point estimates; line segments represent 95% confidence intervals)*



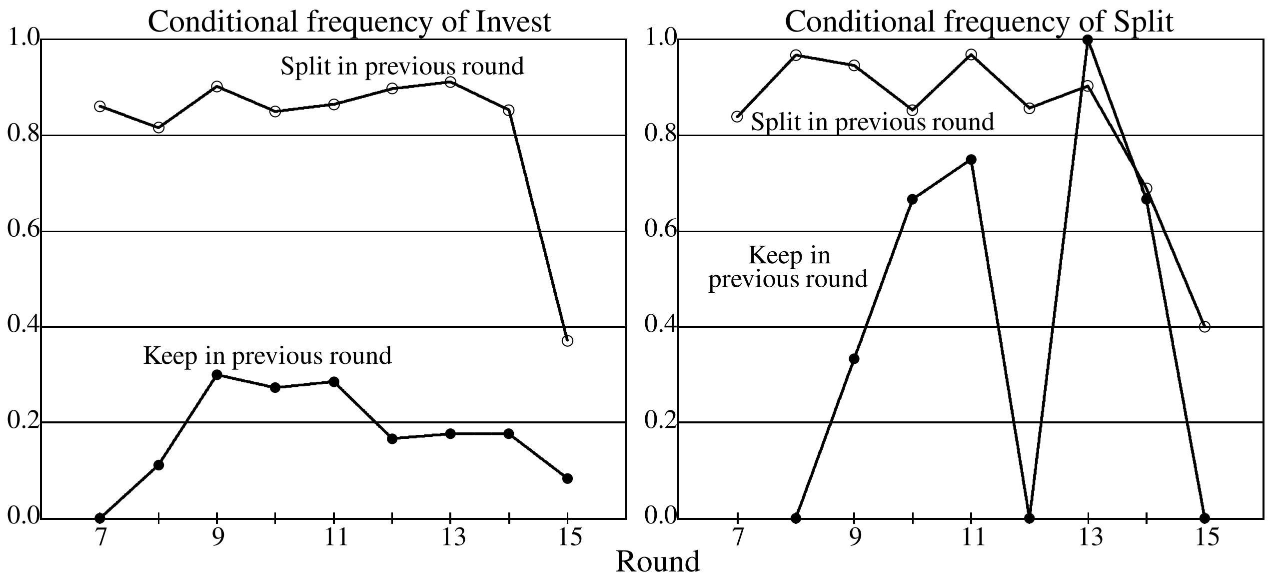
talk message is positive, and even significant in some early rounds, but declines over time and becomes insignificant in later rounds, both when the investor had also observed a Split previous action and when she did not. Comparison of the two series in each panel suggests that interaction effects between messages and observed actions are small enough to be ignored.

4.4. The effects of observed actions

We continue with some descriptive statistics concerning how behavior in a round is associated with the action observed in that round. (Since we have seen that cheap talk has little effect, we leave out a corresponding analysis for messages.) Figure 5 shows that in nearly all rounds, the frequencies of Invest and Split following a Split observed action are higher—and usually substantially so—than following a Keep observed action.[15] Overall, investors chose Invest 82.8% of the time when the observed action was Split, significantly more than the 17.9% of the time they did so after a Keep observed action (Wilcoxon signed–ranks test,

[15]Note that sample sizes are small in most rounds for the case of Split following a Keep observed action.

Figure 5: Subject behavior, disaggregated by observed action (all sessions with observation)



pooled D and WD session–level data, $p \approx 0.062$). Allocators chose Split 85.2% of the time after a Split observed action and 52.4% of the time after a Keep observed action; this difference is also significant ($p \approx 0.016$).

For allocators, the relationship between observed and current–round actions probably just tells us that their choices are positively autocorrelated; we are not arguing that they actually learn anything from their own previous actions. For investors, on the other hand, this correlation suggests a rational reaction to the likelihood that the allocator's past action is a signal of his current action. Alternatively, it could be that investors are motivated by indirect reciprocity, rewarding allocators for past trustworthiness to others by behaving cooperatively now. The sharp dropoff in Invest choices in the last round—when reciprocity should still apply, but allocators' past actions are likely not indicative of their current actions—implies that investors are not blindly reciprocating, but the fact that the dropoff doesn't go all the way to zero suggests that some reciprocity may be in effect (or perhaps that some investors fail to understand how allocators' incentives change in the final round).

4.5. *Summary and discussion*

The positive impact of observation seen in our results warrants some comment, especially when one recalls that it was predicted to have no effect. We suggest two possible explanations, both of which relax the assumptions usually made by game theorists. We stress that these are post hoc explanations, that our experiment cannot distinguish between them, and that still other explanations may be possible. One explanation involves bounded rationality. Solving a single trust game requires only money–payment maximization by the investor and the allocator, and the investor's understanding that the allocator maximizes money payments—that is, one "level of reasoning" by the allocator and two by the investor.[16] Due to our random–matching protocol, the sequence of trust games played in rounds 6–15 of our C treatment can be solved by treating it as ten separate games: there is no connection between the game played in one round and the game played in any other round. The solution therefore still requires only one level of reasoning by the allocator and two by the investor. When cheap talk prior to the investor's decision is added, the extensive form of a single game becomes a bit more complicated, but not much more so. The additional stage has no effect on the subgame perfect equilibrium actions, and equilibrium does not pin down the allocator's message, so the solution of the game with cheap talk still involves the same depth of reasoning.[17]

Adding observation, by contrast, complicates the extensive form substantially.

[16]For more discussion of levels of reasoning in games solvable by backward induction or iterated dominance, see Stahl (1993) or Nagel (1995).

[17]This explanation would continue to apply to unstructured but anonymous forms of cheap talk, such as Bochet et al.'s (2005) "chat room" communication, and as a result fails to explain the increase in cooperation often observed in these treatments. On the other hand, forms of cheap talk that are not anonymous, such as face–to–face communication, allow for the possibility of repeated–game effects, even when subjects are only matched to each other once in a session. (For example, they might meet in the hallway after the session.) As a result, the above argument does not apply, so that it is not necessarily inconsistent with the increases in cooperation typically seen in treatments with face–to–face communication.

Except in round 15, the allocator's Keep/Split decision in the current round will be observed in the next round, so his choice may depend on how he thinks it will affect investors' decisions in subsequent rounds. This connection between games played in different rounds means that it is no longer appropriate to solve each round of the game independently; the solution now involves all of rounds 6–15. It is true that backward induction will lead to the same solution as before, but the closer to the beginning of this sequence of stage games, the more levels of reasoning are required to yield the solution. There is a plethora of experimental evidence that subjects typically use only a small number of levels of reasoning (Selten and Stoecker (1986), Nagel (1995)), so increasing the amount needed to solve the game by even a few levels may lead to a drastically reduced frequency of equilibrium behavior—even if players are still monetary–payment maximizers.

A second (not mutually exclusive to the first) explanation is based on the work of Kreps et al. (1982), who show that if at least a small fraction of players prefers to behave cooperatively in a one–shot social dilemma, then it can pay for all players to cooperate in all but the last few rounds of the repeated game. In the trust game, for example, if it is common knowledge that even a small number of allocators prefer Split over Keep (perhaps due to social preferences), then in early rounds, money–maximizing allocators may prefer to mimic their cooperative brethren—attempting to establish a reputation for cooperative behavior—if this has an effect on investors' subsequent behavior. In Kreps et al.'s model (as well as that of Anderhub et al. (2002), who apply it theoretically and experimentally to the trust game), players play repeatedly against the same opponent, so cooperative behavior in one round is always observed by one's opponent in the next round. In our experiment, allocator choices can affect future opponents' behavior in the D and WD games (where there is observation), but not in the W and basic trust games. As frequencies of Split are nearly always strictly positive—even in the final round—there indeed seems to be at least a small proportion of allocators who prefer Split despite the monetary incentive to choose Keep. Assuming that investors and the remaining allocators understand this, the resulting prediction for the W and basic games would still be low levels of cooperative behavior, but for the D and WD games, the prediction would be high levels of cooperative behavior in early rounds and a dropoff at the end, when the incentive for money–maximizing allocators to mimic cooperators disappears. These predictions conform closely to what we saw in the experiment.[18]

While observation leads to higher levels of cooperation in our experiment, the other side of the coin is that cheap talk performs quite badly. As mentioned in Section 2, the experimental literature suggests that the effect of cheap talk in social dilemmas depends on how it is implemented: unstructured communication usually improves outcomes substantially, while allowing only anonymous, binary messages leads to little if any improvement. Our results are consistent with this pattern, as our addition of anonymous, binary messages to the trust game has little effect, either when observation is possible or when it is not. Future work might examine other implementations of cheap talk, such as unstructured "chat room" communication via computers, or face–to–face communication, in an attempt to give cheap talk its best possible chance of improving outcomes.[19]

We might wonder why no "crowding out" (see Note 2) was seen. Cooperative behavior in sessions with observation was more frequent than in the basic game, and even in sessions with only cheap talk, while levels of cooperative behavior were no higher than in the basic trust game, they were no lower either. We speculate that crowding out may be sensitive to the way the mechanism was framed in the instructions to the subjects. If the instructions had attempted to make clear

[18] Our results are also consistent with Ambrus and Pathak (2007), who show that under certain conditions, a mixed population containing both selfish money–maximizers and reciprocators can yield dynamics that include declining rates of cooperation and restart effects, even when the game's final period is common knowledge amongst the players.

[19] One could also compare cheap talk with costly signaling. An alternative interpretation of our observation game is as an opportunity for allocators to send a costly Split signal for the next round by choosing Split in the current round. If investors react to Split observed actions by choosing Invest, then such costly signaling can be worthwhile to allocators. Also, Bracht and Feltovich (2008) considered a different type of costly signal in the trust game, with the use of "escrow" treatments in which Split messages were costly if the allocator subsequently reneged, but costless otherwise. We found that sufficiently high costs of reneging led to very high levels of cooperation.

that cheap talk and observation were devices to encourage cooperation, they may well have had the opposite effect (as crowding out predicts). However, as they were written, the instructions presented these mechanisms simply as information investors would receive from allocators. This "information" frame might be different enough from a "mechanism" frame to lead to different behavior.

5. Conclusion

In the trust game, the standard theoretical prediction is for low levels of trust and trustworthiness. Because these are important components of social capital, it is worthwhile to discover ways of modifying the game in order to raise their levels. We examine two kinds of modifications, both involving only information from or about the allocator—cheap talk (costless nonbinding messages sent by the allocator) and observed actions (costless accurate reporting of the allocator's previous action)—received by the investor prior to play. Our experiment comprises a basic trust game with neither type of information, a W treatment with cheap talk only, a D treatment with observed actions only, and a WD treatment with cheap talk and observed actions. Under the usual assumptions, none of these variations should affect behavior.

Consistent with this prediction, we find that cheap talk has nearly no effect. In contrast, observed actions have a strong effect. Allocators are much more likely to behave in a trustworthy way if they know their actions will be observed in the next round, suggesting that observation provides a useful check on opportunistic behavior. Investors, for their part, are more likely to show trusting behavior when allocators' actions will be observed in the next round, and also when matched with an allocator who cooperated in his previous opportunity. Cooperative behavior by both players dips sharply in the last round of play (though not to zero), suggesting that players understand to some extent that the check on opportunistic behavior provided by observation is not present in this last round.

Our results are encouraging from an applications standpoint, as we find that coercion is not required to improve outcomes. The "enforcement" in our observation cells is not performed by a third party, but is entirely voluntary on the part of the players. Additionally, we have seen that even a small amount of information about the allocator's history (here, the most recent action) is sufficient to yield a substantial improvement; the entire history (as in Bolton et al. (2004) and Huck et al. (2006)) is not needed. We conjecture that mechanisms like our observation mechanism have the potential to improve efficiency in the outside world as well as the laboratory; indeed, they might work even better there, if time horizons are longer than the 10 rounds of the observation treatments in our experiment.

While we consider the lack of a requirement of coercion to be an advantage of our observation mechanism, it still suffers from the drawback that transmission of information about previous actions is assumed to be 100% correct, 100% of the time. The mechanism thus likely does need a third party (though without coercive power) to collect and pass on this information. A natural next step is to consider mechanisms that do not require such a high standard: yielding information that is correlated with previous actions, but not perfectly so (along the lines of Keser's (2004) reputation mechanisms, discussed in Section 2). Even if information is not perfectly accurate, we speculate that it could still lead to improved outcomes if it is "accurate enough", but this conjecture, along with how accurate is "accurate enough", should be tested.

References

[1] Ambrus, A., P.A. Pathak, 2007. Cooperation over finite horizons: a theory and experiments. Working paper, Harvard University.

[2] Anderhub, V., D. Engelmann, W. Güth, 2002. An experimental study of the repeated trust game with incomplete information. Journal of Economic Behavior & Organization. 48, 197–216.

[3] Andreoni, J., 1988. Why free ride? Strategies and learning in public goods experiments. Journal of Public Economics. 37, 291–304.

[4] Andreoni, J., H. Varian, 1999. Preplay contracting in the prisoners' dilemma. Proceedings of the National Academy of Sciences. 96, 10933–10938.

[5] Berg, J., J. Dickhaut, K. McCabe, 1995. Trust, reciprocity, and social history. Games and Economic Behavior. 10, 122–142.

[6] Bochet, O., T. Page, L. Putterman, 2006. Communication and punishment in voluntary contribution experiments. Journal of Economic Behavior & Organization. 60, 11–26.

[7] Bolton, G.E., E. Katok, A. Ockenfels, 2004. How effective are electronic reputation mechanisms? An experimental investigation. Management Science. 50, 1587–1602.

[8] Bracht, J., N. Feltovich, 2008. Efficiency in the trust game: an experimental study of precommitment. International Journal of Game Theory. 37, 39–72.

[9] Camerer, C., E. Fehr, 2003. Measuring social norms and preferences using experimental games: a guide for social scientists. In J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr H. Gintis, eds., Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small–Scale Societies, Oxford University Press, Oxford, 55–95.

[10] Coleman, J.S., 1988. Social capital in the creation of human capital. American Journal of Sociology. 94, S95–S120.

[11] Duffy, J., N. Feltovich, 1999. Does observation of others affect learning in strategic environments? An experimental study. International Journal of Game Theory. 28, 131–152.

[12] Duffy, J., N. Feltovich, 2002. Do actions speak louder than words? An experimental comparison of observation and cheap talk. Games and Economic Behavior. 39, 1–27.

[13] Duffy, J., N. Feltovich, 2006. Words, deeds and lies: strategic behaviour in games with multiple signals. Review of Economic Studies. 73, 669–688.

[14] Duffy, J., J. Ochs, 2009. Cooperative behavior and the frequency of social interaction. Forthcoming, Games and Economic Behavior.

[15] Falkinger, J., E. Fehr, S. Gächter, R. Winter-Ebmer, 2000. A simple mechanism for the efficient provision of public goods: experimental evidence. American Economic Review. 90, 247–264.

[16] Fehr, E., U. Fischbacher, 2004. Third–party punishment and social norms. Evolution and Human Behavior. 25, 63–87.

[17] Fehr, E., S. Gächter, 2000. Cooperation and punishment in public goods experiments. American Economic Review. 90, 980–994.

[18] Fehr, E., S. Gächter, G. Kirchsteiger, 1997. Reciprocity as a contract enforcement device: experimental evidence. Econometrica. 65, 833–860.

[19] Feltovich, N., 2005. Critical values for the robust rank–order test. Communications in Statistics—Simulation and Computation. 34, 525–547.

[20] Fischbacher, U., 2007. z–Tree: Zurich toolbox for ready–made economic experiments. Experimental Economics. 10, 171–178.

[21] Fukuyama, F., 1995. Trust. Free Press, New York.

[22] Glaeser, E.L., D.I. Laibson, J.A. Scheinkman, C.L. Soutter, 2000. Measuring trust. Quarterly Journal of Economics. 115, 811–846.

[23] Huck, S., H. Normann, J. Oechssler, 2000. Does information about competitors' actions increase or decrease competition in experimental oligopoly markets?" International Journal of Industrial Organization. 18, 39–57.

[24] Huck, S., G.K. Ruchala, J.-R. Tyran, 2006. Competition fosters trust. University of Copenhagen working paper 06-24.

[25] Isaac, R.M., J.M. Walker, 1988. Communication and free–riding behavior: the voluntary contribution mechanism. Economic Inquiry. 26, 585–608.

[26] Kahneman, D., J. Knetsch, R. Thaler, 1986. Fairness and the assumptions of economics. Journal of Business. 59, S285–S300.

[27] Keser, C., 2004. Trust and reputation building in e–commerce. Working paper, University of Göttingen.

[28] Knack, S., P. Keefer, 1997. Does social capital have an economy payoff? A cross–country investigation. Quarterly Journal of Economics. 112, 1251–1288.

[29] Kreps, D., P., Milgrom, P. Roberts, R. Wilson, 1982. Rational cooperation in the finitely repeated prisoner's dilemma. Journal of Economic Theory. 27, 245–252.

[30] La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny, 1997. Trust in large organizations. American Economic Review Papers and Proceedings. 87, 333–338.

[31] Lazzarini, S.G., G.J. Miller, T.R. Zenger, 2004. Order with some law: complementarity versus substitution of formal and informal arrangements. Journal of Law, Economics, & Organization. 20, 261–298.

[32] Nagel, R., 1995. Unraveling in guessing games: an experimental study. American Economic Review. 85, 1313–1326.

[33] Ostrom, E., 2000. Collective action and the evolution of social norms. Journal of Economic Perspectives. 14, 137–158.

[34] Putnam, R., 1993. Making Democracy Work: Civic Traditions in Modern Italy. Princeton University Press, Princeton, NJ.

[35] Putnam, R., 1999. Bowling Alone. John Wiley, New York.

[36] Riedl, A., J.-R. Tyran, 2005. Tax liability side equivalence in gift–exchange labor markets. Journal of Public Economics. 89, 2369–2382.

[37] Sally, D., 1995. Conversation and cooperation in social dilemmas: a meta–analysis of experiments from 1958 to 1992. Rationality and Society. 7, 58–92.

[38] Selten, R., R. Stoecker, 1986. End behavior in sequences of finite prisoner's dilemma supergames: a learning theory approach. Journal of Economic Behavior & Organization. 7, 47–70.

[39] Siegel, S., N.J. Castellan, Jr., 1988. Nonparametric Statistics for the Behavioral Sciences. McGraw–Hill, New York.

[40] Stahl, D., 1993. The evolution of smart$_n$ players. Games and Economic Behavior. 5, 604–617.

[41] Van Huyck, J.B., R.C. Battalio, M.F. Walters, 1995. Commitment versus discretion in the peasant–dictator game. Games and Economic Behavior. 10, 143–170.

[42] Wilson, R.K., J. Sell, 1997. 'Liar, liar...' Cheap talk and reputation in repeated public goods settings. Journal of Conflict Resolution. 41, 695–717.

[43] The World Bank, 1999. What is social capital? http://go.worldbank.org/C0QTRW4QF0.