ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 689 April 2010

The Benefit of Anonymity in Public Goods Games

David Hugh-Jones
David Reinstein

The Benefit of Anonymity in Public Goods Games

David Hugh-Jones and David Reinstein*

March 25, 2010

Abstract

Previous work has found that in social dilemmas, the selfish always free-ride, while others will cooperate if they expect their peers to do so as well. Outcomes may thus depend on conditional cooperators' beliefs about the number of selfish types. An early round of the game may be played anonymously, so that contributions cannot be traced back to particular individuals. By protecting low contributors from potential sanctions, this encourages selfish types to reveal their true preferences in their play. We offer a simple model illustrating when revelation of types can increase contributions, and when only an *anonymous* game can separate types. As a proof of concept, we run a laboratory experiment involving a two-stage public goods game with an exclusion decision between stages. An anonymous first stage led to significantly higher stage-two cooperation than a revealed first stage, a slower decline across the 15 repetitions, unusually high final-stage contributions relative to previous work, and greater profits. Statistical analysis shows that the anonymous first stage reduced uncertainty about types, and this preserved cooperation and led to greater efficiency. Our results suggest that customs such as anonymous church donations may play an important role in building social trust.

Keywords: signaling, anonymity, public goods, club goods, experiments, social trust, reciprocity
JEL codes: H41, Z12, D82.

1 Introduction

The "public goods" environment is one of the most common settings for economic experiments, and these have been run in many shapes and flavors, involving a wide variety of voluntary contribution mechanisms (VCM's).[1] Some core findings have emerged. When sanctions are infeasible and the game is played for a known finite number of repetitions, contributions are typically well above the equilibrium prediction, but below the Pareto optimal level. Contributions decay but remain positive over repetitions.

Why is this so? A recent set of theories involve heterogeneous preferences: some players are self-interested, while others are reciprocators who will cooperate if they expect enough others to do so (Kreps et al., 2001; Schram, 2000; Ostrom, 2000b). Experimental economists have found considerable evidence for "conditional cooperation" and heterogeneity (Keser and van Winden, 2000; Simpson and Willer, 2008;

*David Hugh-Jones is a post-doctoral researcher at the Max Planck Institute for Economics, Jena. David Reinstein is a lecturer in the Department of Economics at Essex University. We thank Toru Suzuki, Ondrej Rydval, Gerlinde Fellner, Martin Leroch, Ryan Mackay and Henry Bottomley ; seminar participants at the MPI, Hamburg University, the University of Essex, the University of Nottingham, the University of Warwick, the University of Amsterdam, and M-BEES.

[1] See Ostrom (2000a), Ledyard (1993), Isaac James et al. (1994), and Plott and Smith (2008) section 6.1 for surveys. We refer to these, and to our own experiment, as "public goods games," although some of these permit exclusion or have some rivalry.

Fischbacher et al., 2001).[2] Other work argues that intrinsic motivations and reciprocity often help improve outcomes (Akerlof and Kranton, 2005; Fehr and Kirchsteiger, 1997).

With heterogeneous preferences, cooperation may be more likely if reciprocators can identify each other and then coordinate on a favorable equilibrium. When conditional cooperators are unknown to one another, they may not cooperate through fear of being cheated. Therefore, as Ostrom (2000a) puts it, "a core question is how potential cooperators signal one another and design institutions that reinforce rather than destroy conditional cooperation."[3] As shown in Frank et al. (1993) and Harrington Jr (1989), this ability to signal is necessary for cooperative types to persist in an evolutionary framework.[4]

How can conditional cooperators know who they are facing? They may learn something by observing others' play over the course of repeated interactions. However, groups will be tempted to punish or exclude members who reveal themselves as free-riders, and if so, free-riders will face strong incentives to act cooperatively – until a particularly crucial episode of play, when the immediate temptation to defect becomes large. Thus, there is a potential time-inconsistency problem: while excluding those revealed to be free-riders will be tempting *ex post*, committing not to punish non-contributors may lead to better information about players' types, and hence better expected utility *ex ante*.

This commitment not to punish can be ensured by playing the game anonymously, so that the distribution of contributions is observed, but contributions cannot be linked back to any one player. Thus an individual's contribution will have little or no impact on her risk of being punished or excluded, reducing the temptation for a free-rider to pretend to be a conditional cooperator. The total number of conditional cooperators is then revealed; if it is high enough, players' mutual trust will be increased and future play will become more cooperative.

Previous work on voluntarily provided public goods and charitable giving has focused on the disadvantages of anonymity: in theoretical work and in some laboratory and field experiments generosity and pro-social behavior is higher when reputation is at stake (Harbaugh, 1998; Glazer and Konrad, 1996; Milinski et al., 2002; Cooter and Broughman, 2005; Andreoni and Petrie, 2004; Soetevent, 2005; Alpizar et al., 2008).[5] By contrast, in our model anonymous play can *increase* cooperation under specified conditions.[6]

This suggests an explanation for the preservation of anonymity in some public goods environments, which is puzzling in light of the evidence suggesting that anonymity decreases contributions. For example, church donations are often taken anonymously. "Pledge cards", which make donations more visible, appear to increase donations, but these face widespread resistance (Hoge et al., 1996). We argue that church congregations may provide mutual support in times of adversity. Publicly identifying donors might increase the amounts given, but congregations would no longer know whether this was driven by real commitment to the

[2] In our model, however, differences in type need not reflect differences in underlying motivations, but could also come from differences in individuals' material benefits from a public good, in beliefs about the private returns to contributions, or in individuals' options outside the group.

[3] See also Brosig (2002) and Fehr and Schmidt (2006).

[4] Ostrom (2000b) summarizes these results, noting that if there is a noisy signal about a player's type that is more accurate than a random signal, trustworthy types will survive as a substantial proportion of the population.

[5] The advantages of anonymity have been discussed in the literature on principal-agent relationships (Holmstrom, 1999; Acemoglu and Robinson, 2006),and in a legislative context (Prat, 2005; Levy, 2007b; 2007a).

[6] Hugh-Jones and Reinstein (2009) focuses on an anonymous "burning money" signaling game. In the present paper the signaling game takes the same form as the main public goods game. It seems natural that the signaling institution might resemble the basic collective action problem; as good types benefit more from their own contribution, it becomes cheaper for them to signal (as in the standard model of Spence, 1973), and thus easier to separate the types. Our experiment uses this format, with a small first-stage public goods game followed by a larger game.

church, and would then lack the trust to cooperate in difficult periods.[7]

If costly voting can signal public support for a particular policy or party (Londregan and Vindigni, 2006), then anonymous voting may provide a better signal of commitment than other non-anonymous activities. In the 1980s, the UK Conservative government forced unions to ballot their members secretly before a strike, expecting this to reduce the effectiveness of strike threats. In fact, turnout in the ballot provided a costly signal of members' commitment to the strike, and this may have actually helped unions negotiate with management, and increased members' willingness to take action after a "yes" vote (Martin et al., 1991).

Other social institutions can be interpreted as technologies that ensure anonymity – even in small groups where members' behavior could be easily observed. Some forms of ritual music and dance conceal participants' identities and/or effort levels. On a mundane level, contributions to a "coffee kitty" are often anonymous, as are "Secret Santa" gifts to one's co-workers.[8]

In this paper we demonstrate the benefits of anonymity using a simple model. We then report a laboratory experiment testing our theory – a repeated two-stage public goods game, where players can be excluded in between the stages. In our experiment, when the first stage was anonymous, second stage contributions were substantially and consistently higher, and declined little (relative to previous work) over the fifteen repetitions. This unusual result demonstrates that anonymity can increase public goods contributions, and can lead to persistent cooperation even without the threat of punishment. A close analysis of our data suggests this occurred as our model proposes. Anonymity caused players to play more "honestly" in the first stage, by reducing the threat of exclusion. Players in the second stage were then better informed about other group members' likely contributions, and could contribute generously when they expected others to do so too.

The remainder of the paper is organized as follows. In section 2, we develop a simple model to illustrate our idea, and to motivate our hypotheses for the experiment. In section 3 we describe our experiment explain our design choices (3.1), and describe our hypotheses (3.2). Section 3.3 presents our basic results. We conclude in section 4 with an interpretation and motivation of our results, and we offer suggestions for future research.

2 Model: conditional cooperation and the minigame

N players each choose to donate $x_i \geq 0$ to a common good. Let $\bar{X} = \sum_{j=1}^{N} x_j/N$ be average contributions and $\bar{X}_{-i} = \sum_{j \neq i} x_j/(N-1)$ be the average contribution of all players except player i. There are two types of players. Player i's welfare is given by

$$\alpha \bar{X} - x_i + \psi(x_i, \bar{X}_{-i}) \tag{1}$$

[7] Several religious traditions particularly encourage anonymous giving. In Judaism Maimonides' Mishneh Torah, Laws of Gifts to the Poor 10:7-14, specifically ranks double-blind anonymous charitable giving above recognized giving. Jesus' admonition that "when you do some act of charity, let not your right hand know what your left hand is doing" (Matthew 6.3) has been interpreted as an injunction against publicizing one's philanthropy, and this is echoed in the Koran (section 2.271). Note that we do not claim that our model is the *only* reason that religious groups encourage anonymous giving. For example, church donations may be kept secret to avoid embarrasing poorer congregants. Instead, our model points out an *additional* potential function of the anonymity.

[8] In (Hugh-Jones and Reinstein, 2009), we develop these examples in more detail, and offer others in which the anonymous action appears to have no independent value (resembling a "burnt money" signal).

if she is a good type, and

$$\alpha\bar{X} - x_i \tag{2}$$

if she is a bad type. Player types are independent: the probability of a good type is $\pi \in (0,1)$. $\alpha \in (1,N)$ is the multiplier for the unconditional benefit of the good (we will refer to $\alpha\bar{X} - x_i$ as "material welfare"). Write $\bar{\alpha} = \alpha/N$ for the marginal unconditional benefit of one's own donations. ψ represents a (psychological or material) benefit received by good types only, which is twice differentiable, strictly concave and strictly increasing in both its arguments, with a positive cross partial, and $\psi(0,0) = 0$.We refer to the ψ component as the CC (conditional cooperation) payoffs. The positive cross-partial can be seen as reflecting a reciprocity motive: good types wish to donate more when others contribute more.

To ensure the existence of equilibrium, we assume that $\psi_1(0,0) > 1 - \bar{\alpha}$ and that ψ_1 and ψ_2 are bounded, with $\psi_1(x,X) \to 0$ as $x \to \infty$ for all X; and that there is some finite $\bar{x}$ with $\psi_1(\bar{x},X) < 1 - \bar{\alpha}$ for all X. We also assume that $\frac{\psi_{12}(x,X)}{\psi_{11}(x,X)} \in (-k\frac{x}{X}, 0)$ for $X > 0$ and some fixed $k \in (0,1)$ (i.e., that the marginal benefit of one's own contribution is "not too sensitive" to others' contributions). This technical condition guarantees a unique equilibrium.

A bad player never contributes anything, since $\bar{\alpha} < 1$. A good player who expects that others' average donations will be exactly X solves the first order condition and gives x_i such that[9]

$$\psi_1(x_i, X) = 1 - \bar{\alpha}. \tag{3}$$

Write $b(X)$ for this best response, $b(X) = x_i$ satisfying (3). This is single-valued by the strict concavity of ψ.

When others' donations are uncertain, good types' optimal donations satisfy

$$E_{\bar{X}_{-i}} \psi_1(x_i, \bar{X}_{-i}) = 1 - \bar{\alpha}. \tag{4}$$

Suppose that, prior to the main game, there are revealed to be $g+1$ good players in total, with the identity of these players either common knowledge, or completely unknown. Then, as we prove in the Appendix, there is a unique equilibrium in which all good types donate the same amount $x_g > 0$. On the other hand, suppose that players only know the "prior" distribution of good players. As we show in the Appendix, this again implies a unique symmetric equilibrium, in which all good types contribute $x^* > 0$.

As our anonymous minigame institution is designed to reveal the number of good types, we would like to know the conditions under which common knowledge of types will increase donations on average. This will hold when

$$\sum_{g=0}^{N-1} \Pi(g)(g+1)x^* < \sum_{g=0}^{N-1} \Pi(g)(g+1)x_g \tag{5}$$

which need not hold in general. The next Lemma gives a sufficient condition.

Lemma 1. *Common knowledge of the number of good types increases donations ex ante when $\psi_1(x,X)$ is weakly concave in X and $b(\cdot)$ is weakly convex.*[10]

[9] ψ_l is the derivative with respect to the l'th argument.

[10] The weak concavity of ψ_1 can be thought of as making good types "risk averse" in terms of how much they contribute, hence knowledge will increase contributions. The condition on $b(\cdot)$ is simply a condition on the locus of points (x,X) such that

If $\psi(x,X)$ reflects an individual's social preferences (rather than a profit function), we cannot directly examine it, so we cannot *a priori* know that contributions will be higher when the number of good types is known. However, existing experimental work (cited in section 1) suggests that people's uncertainty about others' contributions depresses their own contributions. We adopt this as a working hypothesis for our experiment.

2.1 The first round (minigame)

When common knowledge of types leads to increased donations, this is likely to be advantageous to a group. However, players cannot simply be asked their type: bad types would claim to be good so as to increase others' donations. Instead, players must give some costly signal of their type. This might occur, for example, if there are multiple rounds of the public goods game. Then, contributions in early rounds may signal type and inform players what to expect in later rounds.

The division into rounds might be a deliberate choice by the group: for instance, if church donations are taken regularly every week. Or it may occur naturally in environments that require ongoing cooperation, such as the South Asian community irrigation systems described in Bardhan (2000). In either case, we assume that there is a crucial "last round," where the stakes are particularly high.. This can be thought of as an emergency, where the existence of the group may be threatened, or, generalizing this, where the costs and benefits of immediate play outweigh the "shadow of the future".

In our model, there are just 2 rounds. We call the first round a "minigame". In between the rounds, players may be excluded from the second round, in which case they receive zero utility from it. We assume that the first round of contributions is simply a scaled-down version of the main collective action problem, in which both material and CC welfare is multiplied by $D < 1$.[11] Thus, i's total welfare, using superscripts of 1 and 2 for rounds 1 and 2 respectively, is

$$D\left[\alpha\bar{X}^1 - x_i^1 + \tau_i\psi(x_i^1, \bar{X}_{-i}^1)\right] + \eta_i\left[\alpha\bar{X}^2 - x_i^2 + \tau_i\psi(x_i^2, \bar{X}_{-i}^2)\right]$$

where $\tau_i = 1$ if i is good and 0 otherwise, and $\eta_i = 1$ if player i is included in the second round, $\eta_i = 0$ otherwise.

There are two different types of minigame. After a *revealed* minigame, all players' contributions are public knowledge. After an *anonymous* minigame, only the profile of contributions is revealed, so exclusion cannot be targeted at any particular player based on her round 1 contribution.

For technical simplicity, we assume that exclusions are implemented only when they will increase the proportion of good types in the second round. Thus, in the revealed institution, only good types are included. In the anonymous institution, where bad individuals cannot be targeted for exclusion, all players are included. However, proposition 2 will hold even if we allow a certain proportion to be randomly excluded under anonymity. [12]

$\psi_1(x,X) = 1-\bar{\alpha}$. Hence this is equivalent to a condition on the model's primals. As an example, the conditions of the Lemma hold if $\psi(x,X) = x^\gamma X^{1-\gamma}$ with $\gamma \in (0,1)$.

[11] Other assumptions are possible: for example, first round material welfare could be the same but with smaller maximum contributions, $x_i \in [0,D]$, with the CC welfare function unchanged as $\psi(x,X)$. Or ψ could be defined over both rounds of contributions simultaneously, to allow payoffs from conditional cooperation over different rounds. Our formulation is chosen for simplicity, but our results are not sensitive to this assumption, since our proofs do not use the fact that ψ is identical between rounds.

[12] A wide range of assumptions about the exclusion mechanism are possible – it could maximize aggregate welfare, with or

We look for a separating equilibrium, where the types play differently in the minigame. When D is very small, the incentive to pretend to be good and thus increase round 2 contributions will dominate the incentive to contribute little in the minigame, even for bad type players, and separation will be impossible. When D is larger, a separating equilibrium will be possible. An anonymous minigame allows separating equilibria for lower values of D. The reason is intuitive: when play is anonymous, the cost of playing selfishly is only that others donate less in the second round. When play is revealed, selfish play results in exclusion from the second round. As a result, the incentive to pool is higher in the revealed institution.

We refine our set of equilibria using the Intuitive Criterion. This yields clearer results, and the restriction imposed is indeed intuitive here: it requires that in equilibrium, good types cannot profitably deviate towards the donation that they prefer in the stage-game, unless a bad type can also profit from making such a deviation.

Proposition 2. *In the anonymous minigame, there is an Intuitive separating equilibrium (only) for values of D above a fixed $\hat{D}$. In the revealed minigame, there is an Intuitive separating equilibrium (only) for values of D above D^*, where $D^* > \hat{D}$.*

If $D \in (\hat{D}, D^*)$, a separating equilibrium is possible in an anonymous minigame but not in a revealed minigame. In these equilibria, good types donate positively in the first round, while bad types donate 0. These ideas motivate our experimental hypotheses.

3 Experiment

We test our model and its predictions in an ambiguous, but well-studied environment, the public goods game. We set up an environment with a signaling institution of a fixed size, and observe how players learn to make use of it. If they can do so during a brief laboratory experiment, this suggests that they could also do so in the field. Similarly, if signaling institutions without anonymity fail in the way we expect, real-world institutions may also face problems of pandering.

We do not induce heterogeneity in preferences, instead relying on subjects' "homegrown values" (Harrison, 2002). Behavior with induced values may lack external validity. For example, the play of selfish players who are assigned material payoffs to mimic the preferences of hypothetical conditional cooperators may not be a good predictor of the behavior of real-world conditional cooperators.

While we recognize the case for a stark experimental design to test simple hypotheses, these tests would be redundant (repetitive of previous work) or misleading here. As described in section 1, previous experiments have already found evidence for many of the components of our model; but these components may be context-specific: for example, we do not know if conditionally cooperative subjects can infer whether others are conditionally cooperative from their behavior.[13] Although we leave many "free parameters" and test

without CC utility, or the welfare of good types, or average donations, or the difference between good and bad type welfare. Or exclusions could be modelled game-theoretically, for instance as the result of majority vote. In all these cases, it will be beneficial to exclude bad types and since revealed institutions allow targeting of bad types, the revealed institution will increase the incentive to pool with good types. (Also, since exclusion is more effective in the revealed institution, it is likely to be used more, which will again increase the incentive to pool.) Thus, we conjecture that our results can be extended to a wide set of exclusion mechanisms.

[13] Furthermore, we do not know whether individuals can predict their own future preferences, nor whether they are strategically naive or sophisticated over "behavioral" preferences. We do not know whether the relevant individuals are able to make reasonable Bayesian inferences from others behavior, nor whether their priors over the distribution of types are consistent. Even if subjects

several elements at once, our proof-of-concept experiment demonstrates that the predictions of our model hold in at least one environment.

3.1 Design and implementation

We ran five sessions on a total of 120 subjects from the standard pool at the University of Jena. The experiment lasted approximately one hour. Subjects were paid a show-up of 2.50 Euros in addition to the profits mentioned below. Our design is as follows. Thirty subjects enter the session; fifteen are randomly assigned to the *anonymous*, and fifteen to the *revealed* treatment. Subjects play 15 repetitions[14], and always remain within the same anonymity treatment, i.e., this is a between-subjects design. For each repetition, subjects are randomly assigned to groups of five – we use a "stranger matching" design. A repetition consists of two linear public goods games. Players in each group are randomly numbered from 1 to 5. In stage one of this game, the *signaling institution*, players 1-3 (the "leaders")[15] play a smaller-stakes VCM game among themselves. Players 4-5 (the "followers") then observe the leaders' contributions and may choose to "exclude" one, or none, of the leaders from the second stage. One exclusion decision is implemented, at random. In the revealed treatment, followers observe the amount each leader contributed along with her player number, and can exclude on this basis. In the anonymous treatment, leader contributions are not linked to player numbers, and hence followers cannot target specific leaders for exclusion.[16] This is the only difference between treatments.

Our discussion and models assume all players are present in the minigame. By separating leaders, who participate in the first stage, from followers who vote on exclusion, we ensure that exclusion decisions were not motivated by direct reciprocity or revenge motivations arising out of stage one.

In the second public goods game, all players make contribution decisions, but an excluded player's decision is ignored in computing payoffs.[17] We do not announce the exclusion decision until the end of a repetition. This allows us an additional data point per round, and ensures that second-round contribution decisions are made in a relatively homogeneous environment (across treatments and repetitions), as subjects can not yet be certain whether an exclusion has been made. Finally, all players learn choices, profits, and exclusion decisions. The overall structure of the experiment can be seen in figure 1.

[Figure 1 about here]

Previous experiments have found that a greater marginal per capita return (MPCR) tends to increase contributions, as does a larger group size, holding the MPCR constant (hence increasing total returns). However, if the total return rate is kept constant as group size increases, the effect of the decrease in the MPCR dominates, and contributions decline (Ledyard, 1993). It is impossible to keep both MPCR and

make inferences as we predict and play strategically, there is the issue of coordination on an equilibrium. Thus, we simultaneously test for conditionally cooperative preferences, measure the nature and extent of these preferences, and examine how subjects infer others' likely play and respond strategically.

[14] Six repetitions in the pilot version.

[15] To avoid an experimenter-demand effect, the terms "leader" and "follower" were not used in the experiment itself. In even repetitions, first players 1-2 were selected for each group, from a pool made up of the previous repetition's players 4-5; then players 3-5 were selected for each group from the remaining players. This ensured a reasonable balance of leader/follower roles across subjects.

[16] In the anonymous treatment the choice is essentially whether or not to exclude a randomly chosen leader.

[17] If a player is excluded the returns from the public good are calculated based on the contributions of the remaining four players and shared among them. The excluded player simply receives her initial second-stage endowment.

total return rate the same when a player is excluded. We set our total return rates (of 1.5 and 2 in stage 1 and 2 respectively) to have the same MPCR for the first and second stage in the *presence* of exclusion ($\frac{1.5}{3} = \frac{2}{4} = 0.5$). If there is no exclusion the MPCR is slightly lower in stage 2 ($\frac{2}{5} = 0.4$).[18]

We argue that this is not driving our results. Firstly, the difference is small (0.4 versus 0.5), and subjects do not know whether an exclusion will take place when they are making their decision, so that the potential *expected* difference between treatments is even smaller. Second, an exclusion is slightly more likely in the revealed case. If subjects are aware of this then the expected second stage MPCR is higher in the revealed treatment, and this would presumably lead to *greater* contributions in the revealed treatment. Thus, the predicted bias works in the opposite direction of our finding.

In repetitions 3, 7, 11 and 15, we also elicited incentivized guesses about other players' second stage contributions, using a quadratic scoring rule. At the end of the game, participants received their payoffs from two randomly chosen repetitions (one for the first stage, one for the second stage) and for their guesses.[19]

3.2 Hypotheses

We expect a range of types, rather than the model's two types. Nevertheless, in equilibrium, some or all types may pool in the first round. We expect that our chosen payoff sizes cannot sustain full separation when contributions are revealed, but can separate types to a larger extent when they are anonymous. Motivated by previous experimental evidence, and intuition derived from our model, we make the following hypotheses:

1. The correlation between an individual's stage 1 and stage 2 contributions is non-negative and greater in the anonymous treatment. *Intuition:* With less than complete pooling in the anonymous case, the stage 1 investment will be informative about stage 2 investment. More pooling will be observed in the revealed treatment as a player's contribution has a greater impact on the risk of exclusion (see Hypothesis 3). Invoking our stylized model, if the conditions for proposition 1 hold, there will be separation in the anonymous treatment only.

2. Players exclude less given higher stage 1 contributions. *Intuition:* All other players benefit from excluding a player who donates less than average. If the stage 1 contributions are informative of stage 2 choices, the (expected) material incentive to exclude a player will decrease in that player's stage 1 contribution. The same effect would be caused by a fairness motive, as players will prefer to punish uncooperative players.

3. Lowering one's contribution will increase the risk of being excluded more in the revealed than in the anonymous treatment.[20] *Intuition:* This hypothesis stems from the simple fact that under anonymity a leader cannot be targeted based on her contribution. Her decision therefore only affects the overall probability of an exclusion. If there is an exclusion, she will only be "hit" with 1/3 probability. Hence

[18] We could have let the total return rate change in the presence of exclusion to preserve the MPCR, but this would have been very difficult to explain to the subjects.

[19] We provide screen shots of key stages in the online appendix.

[20] Note that we allow that there may be *some* exclusion in the anonymous treatment. Unlike in our simple model, the leaders are a subset of the full group (for that repetition). Thus first-stage play may reveal that the leaders' average type is worse than the population average (prior belief), implying that randomly excluding a leader could be expected to increase the share of good types included in the second stage.

this hypothesis will hold unless the impact of an action on the probability of an exclusion is at least three times higher in the anonymous treatment, which seems unlikely.

4. Players' subjective expectations of stage 2 contributions respond more to stage 1 contributions in the anonymous treatment than in the revealed treatment. *Intuition:* Players anticipate Hypothesis 1, and their expectations reflect this.

5. Players' hypotheses of leaders' stage 2 contributions are more informative (i.e., explain a greater share of the variation in actual stage 2 contributions) in the anonymous treatment than in the revealed treatment. *Intuition*: As explained above, stage 1 contributions are likely to have more information content (about true preference and thus likely stage 2 behavior) in the anonymous case then in the revealed case.

6. Players' stage 2 contributions increase with their expectations of others' stage 2 contributions. *Intuition:* This reflects conditionally cooperative preferences.

7. Under anonymity, players' stage 2 contributions increase in others' stage 1 contributions. *Intuition*: As argued above, there will be some "separation" in the anonymous case, and thus behavior in both stage 1 and stage 2 will reflect a player's social preferences and beliefs.

8. The extra information in the anonymous treatment results in higher contributions. *Intuition:* This is not inevitable in our model (see Lemma 1), but since public goods experiments have consistently found that contributions decline over repetitions,[21] we suspect that the equilibrium without a successful signaling institution will have low cooperation levels, so that signaling should be an improvement.

3.3 Results

Overview, overall contributions (Hypothesis 8)

[Table 1 about here]

Summary statistics are shown in Table 1. For both treatments, stage 1 investments are within the range typically found in prior work (Ledyard, 1993). Exclusion was common in both treatments, but subjects in the revealed treatment were significantly more likely to vote to exclude someone.[22] Although predictions for others' stage 2 investments were lower in the revealed case, they still were somewhat overoptimistic. The final column pertains to the number of votes to exclude a particular leader; a leader's overall probability of being excluded was roughly 15%. [23]

[Figures 2 *and 3 about here]*

[21] In explaining this pattern, Ostrom (2000b) argues that conditional cooperators (good types) begin optimistic, and some "egoists" strategically pool with them (as in Kreps et al. (2001)). (As Holt and Laury (2009) note, the former group must "systematically overestimate" their prevalence). As the end of the game approaches and free riding occurs, the good types become disappointed, reducing their contributions and discouraging other good types from contributing. "Without ... institutional mechanisms to stop the downward cascade, eventually only the most determined conditional cooperators continue to make positive contributions in the final rounds." We see our anonymous minigame as one such mechanism.

[22] This difference is significant in a 2-tailed t-test (p=0.006) and in Fisher's exact test (p=0.007).

[23] In the anonymous case, since the selection of whom to exclude is effectively random, we replace the actual number of votes against a player with one third of the total votes (0,1,or 2) to exclude in the relevant repetition and group. This substitution reduces random noise but our results are not sensitive to this substitution.

As Figure 2 demonstrates, stage 1 investments remained fairly constant across repetitions in both treatments. In contrast, stage 2 investments began higher and declined much less under the anonymous treatment, remaining at above 30% on average while falling to around 10% in the revealed treatment. We argue that the anonymous first stage made this persistent cooperation – which is unusual in the absence of punishment (Ledyard, 1993) – possible. Figure 3 shows that these patterns are similar across sessions.[24] Stage 2 average investment in the final stage is strictly lower in the revealed case for *all* sessions. The difference in average stage 2 investments between treatments was statistically significant in both parametric and nonparametric tests (Wilcoxon rank-sum $p = 0.07$, taking the treatment/session as the unit of observation). Anonymity was materially beneficial to subjects. Average net profits (earnings over both stages less endowments) were 4.58 Euros in the anonymous treatment and 3.35 Euros in the revealed treatment.[25] This result supports Hypothesis 8.

Correlation between stages (Hypothesis 1)

[Figure 4 about here]

Figure 4 shows frequencies of stage 2 contributions by stage 1 contributions for each treatment (bubble width indicates number of observations), for the later repetitions, presumably after some strategic learning has taken place. Hypothesis 1 appears to hold: there is a positive correlation between giving in the stages, and the correlation is much stronger in the anonymous treatment. In particular the revealed treatment shows many high stage 1 contributions followed by low stage 2 contributions, which suggests attempts to avoid exclusion. Next we decompose the variance into its explained and unexplained components, reporting marginal and total sums of squares.[26]

[Table 2 about here]

A subjects' first stage investment explains much of the variance her stage 2 investment in the anonymous treatment, while in the revealed treatment it explains almost nothing. It is not just the *presence* of an anonymous stage 1 contribution that matters, but also its magnitude; a 1 ecu investment explains little, while larger investments matter a great deal.[27]

Almost all the explanatory power of first stage investment is via individual heterogeneity. In the final two columns, after conditioning on subject-specific effects, first stage investment explains little of the remaining variation for either treatment. That is, first stage contributions "explain" second stage contributions in the anonymous treatment, because they are reliable signals of the individual leaders' types.[28]

[24] note the pilot session involved only 5 repetitions for each treatment

[25] This difference is strongly significant in t-tests and rank sum tests at the group-repetition level (p=0.000 for both tests). Figure 7 in the Online Appendix plots these profits across repetitions for each treatment.

[26] Interpreting this in a regression framework, the marginal sum of squares can be interpreted as "the reduction in R-sq if you removed that variable only." These add up to the TSS only if the variables are exactly orthogonal.

[27] Regression analysis of stage 2 contributions (see table 7 in the Online Appendix and accompanying notes) yielded comparable results.

[28] Nonetheless, stage 1 investment is obviously an important signal for conditionally cooperative subjects. As subjects do not know the identity of the others in their group, the subject-fixed effects are not directly observable to the subjects themselves – hence they must rely on stage 1 investment as a measure of the other subjects' types.

Aside: Econometric Discussion

While observations at the treatment/session level are strictly independent, they do not fully exploit the information in the data. Our design uses "imperfect" stranger matching; subjects play many repetitions in a limited pool. As in all such experiments the per-subject observations are not completely independent, and play may be affected by experience in earlier repetitions. To deal with this, we estimate robust standard errors, clustered at either the subject level or the treatment/session level as appropriate. Where noted, we also use a set of four control variables for subject *i*'s experience: (i) the average stage 1 investment of other subjects in *i*'s group for the previous repetition, (ii) the same for stage 2, and (iii,iv) the means of each of these over *all* of *i*'s previous repetitions.[29]

Exclusion (Hypotheses 2 and 3)

Table 3 gives Probit regressions for a follower's decision whether to exclude any leaders. As Hypothesis 2 implies, in both treatments the probability of an exclusion decreases in the minimum stage 1 contribution. For *early* repetitions, the minimum gift has a negative and sometimes significant impact on the probability of an exclusion for both treatments. However, for the revealed treatment the effect is significantly greater and persists even through the later repetitions.[30]

[Tables 3 and 4 about here]

To test Hypothesis 3, we examine the impact of a leader's gift on the probability that she is excluded. Table 4 demonstrates that the probability a leader was excluded[31] varied inversely with her stage 1 investment, and this effect was much stronger in the revealed treatment. The overall conditional probabilities of exclusion were 78%, 53%, 12%, 6%, and 4% given revealed stage 1 contributions of 0,1,2,3, and 4 respectively, with an even steeper slope in later stages (probabilities are imputed as one third the total "simulated" votes against a subject).

Beliefs (Hypotheses 4 and 5)

[Figures 5 *and* 6 *about here]*

[Table 5 about here]

Figures 5 and 6 show the density of players' predictions of leaders' stage two contributions, conditioned on the leaders' actual stage 1 contributions. Overall, in both treatments the modal stage 2 prediction is about twice the actual stage 1 contribution. However, in later repetitions, players in the revealed treatment became more skeptical: predicted contributions become much lower for high stage 1 contributors.

Table 5 regresses players' predictions about leaders' stage 2 contributions on the leaders' actual stage 1 contributions.[32] In line with Hypothesis 5, the coefficient of first-stage investment is somewhat lower

[29]This specification was chosen for parsimony and based on some preliminary tests. The simple lag term is motivated by the idea that memory has a recency bias. Regressions allowing an intercept for sessions/treatments yielded similar results.

[30]In some regressions anonymous median gifts have a negative and significant coefficient, while the coefficient on revealed median gifts is positive and significant. We speculate that this is because greater contrast between the contributions makes the exclusion decision harder in the former case and easier in the latter.

[31]These probabilities are simulated in the anonymous case to reduce random noise; see footnote 23.

[32]In the anonymous treatment predictions were for (e.g.) "the player who contributed 4 ecus."

(significantly so in columns 1 and 5) in the revealed case, although the summed coefficient remains significant. As predicted by Hypothesis 5, subjects' guesses were significantly better in the anonymous than in the revealed treatment, with correlations to targets' choices of 0.39 and 0.26 respectively (see Table 8 in the online appendix and accompanying notes).

Conditional cooperation (Hypotheses 6 and 7)

We expect a player's second stage investments to increase in her expectation of others' investments. Because the prediction itself may be correlated to subject-specific unobservables (e.g., more generous people may be more optimistic about others)[33] we control for a subject-specific effect.[34]

[Table 6 about here]

The first column of Table 6 measures the relationship between a player's second stage investment and her expectations of others' contributions, in other words, it measures her level of conditional cooperation. We only include followers, to rule out motives such as direct reciprocity for stage 1 or bitterness from the perceived probability of being excluded. We allow the slope in the minimum guess to vary by treatment, as minimum givers in the revealed treatment are likely to be excluded. As before, these regressions control for a per-treatment time trend; here we also include a "final repetition" dummy to allow for an end-game effect. There is clear evidence of conditional cooperation, in line with Hypothesis 6: the coefficients on each of the guesses (the lowest, middle, and highest guesses for leader subjects' stage 2 investments, and the guess for the other follower subject) are positive, and these are jointly significant in an F-test at the 5% level.[35]

Confirming Hypothesis 7, players' stage 2 investments increase in others' *anonymous* first stage investments. As noted above, the probability of exclusion is nonlinear in first stage investment; hence subjects' inferences may also be nonlinear. In columns 2 and 3 we use the number of leaders who gave 2-4 and 3-4 ecus in the first stage as independent variables.[36] Column 2 shows that these donations had strong effects. As column 3 shows, these persist into the later repetitions for the anonymous treatment, but disappear in the revealed case.[37] This suggests that followers grow less confident in revealed first-stage investments as predictors of leaders' second-stage behavior, and thus cease to respond positively.[38]

[33] From Ledyard, 1993, citing Orbell et al. (1988): "One of our most consistent findings throughout these studies – a finding replicated by others' work – is that cooperators expect significantly more cooperation than do defectors. This result has been found both when payoffs are "step-level" (when contributions from a subset of k subjects ensure provision of a benefit to all) and when they are "symmetric" (when all contributions ensure a constant benefit to all)."

[34] We do not include lagged controls for a subject's experience in previous repetitions here, as we expect the effect of these to be subsumed in the subjects' expectations; the results are not sensitive to this.

[35] This interpretation might be criticized on the grounds that the subject may first choose how much to invest and her prediction may be an ex-post rationalization of this choice (see Fehr and Schmidt 2006). In response we first note that our subjects' guesses are financially motivated. We second point to our evidence below that followers also respond to stage 1 *contributions*. Finally, instrumental variables regressions (available in the online appendix) using others' first stage investments as instruments for predicted stage 2 contributions strongly support our results (the instrumented '*average guess*' is positive and strongly significant).

[36] Since the leader who gives the least is likely to be excluded in the revealed treatment, we do not include the lowest gift in this count. Hence these variables may take values 0,1, or 2.

[37] The summed coefficients for the revealed case are significantly different from the anonymous case, but not significantly different from 0.

[38] It might be argued that the slightly higher probability that a leader is excluded in a revealed repetition (47% versus 42%), or the fact that this exclusion is more personally targeted, could be driving the greater decline in stage 2 contributions. Subjects who have been excluded in previous stages might become embittered and thus contribute less. As we show in the online appendix (table 9 and comments), the data does not support this hypothesis, and our results are robust to the inclusion of a control for "previous exclusion."

These regressions are "reduced form": in our model, first stage contributions only affect stage 2 contributions via their effect on expectations of stage 2 contributions. Column 4 tests this interpretation by regressing followers' investments on both their guesses and the leader investment counts.[39] The coefficients on the guesses themselves remain largely positive and are jointly strongly significant (in an F-test). However, after controlling for guesses, the coefficients on the leaders' investment variables are smaller and are no longer positive and significant.

4 Conclusion

The experimental results are consistent with our hypotheses. The anonymous first round appears to have helped players learn about the preferences of their fellow group members, and this decrease in uncertainty increased efficiency. Second stage contributions were significantly and substantially higher in the anonymous treatment, as were overall profits. The data also support our account of the mechanism behind this: leaders contributed to avoid exclusion under the revealed treatment; thus contributions were more closely correlated between the stages under anonymity, and as a result beliefs were more accurate. In early repetitions followers' gifts varied positively with leaders' gifts; this relationship persisted into later repetitions only in the anonymous case, disappearing in the revealed treatment. We claim that in the latter case, good types could no longer confidently identify other good types, and this increasing uncertainty reduced second stage investments. The robust positive relationship between subjects' investments and their predictions of others' investments supports this story.

Until now, anonymity has been seen as endangering public goods provision, since anonymous actors do not need to worry about their reputations. However, some social institutions, such as church donations, deliberately preserve anonymity. Our theory can explain this anomaly: an anonymous public goods institution may allow participants to learn each others' true character, and may thus make subsequent play more socially efficient. Our experiment's results are consistent with this. We hope that our work will motivate future research into the signaling value of particular anonymous institutions in the real world.

The gains from anonymity are likely to be large when "final stage" cooperation is very important: that is, cooperation which cannot be enforced by the threat of subsequent sanctions. For instance, episodes of conflict, natural disasters, and economic crises all put a premium on groups' ability to cooperate, and may simultaneously make the environment uncertain, so that future group interactions cannot be guaranteed. Levels of cooperation are known to vary widely in such situations, which provide an important motivation for the study of laboratory public goods games.[40]

In some lab and field work, institutions that can sustain cooperation while present, such as punishment, lead to *less* cooperation when they are removed (Ostrom, 2000b; Frohlich and Oppenheimer, 1996 Cardenas

[39] We also ran this same regression controlling for a subject fixed effect, for the reasons previously mentioned (available by request). This yielded the same qualitative result but relied on a very small number of observations, since the fixed effect could only be identified for subjects who were followers in multiple prediction repetitions and invested a positive amount in each. Hence this data yielded very little conditional variation in the "Num Ldrs" variables, few degrees of freedom, and very wide standard errors for the corresponding coefficients.

[40] For example, consider the case of blackouts in New York City: "... it is also hard to predict when looting will erupt, and when it won't. New York's 1965 blackout was famous for the citywide bonhomie it produced, as well as for the baby boom nine months later. But July 1977 was different: When the lights went out then, if one lived in Greenwich Village it felt like the big block party was back; but in poor black and Hispanic neighborhoods, hundreds of stores were looted and 25 fires still burned the next morning." <http://www.nytimes.com/2010/03/07/weekinreview/07mcneil.html?pagewanted=2&hpw >

et al., 2000). Our theory suggests an explanation: cooperative behavior can emerge in the context of local institutions and local moral norms, and when it does, it is informative about the pro-social preferences ("type") of the participants. Enforcement institutions can destroy this information, by forcing everyone to behave well. We intend to examine this hypothesis in future work.

Finally, we note that our results have implications for policy. Karlan (2005) finds that trustworthiness in a laboratory trust game (run on Peruvian microcredit participants) predicts repayment of a loan "enforced almost entirely through social pressure." Games like this, with anonymous participation, might help participants build trust (in cases where it is warranted) and lead to greater contributions to lending pools and other collective goods. Future field experiments should explore this possibility.

References

Acemoglu, D. and J. A. Robinson (2006). *Economic origins of dictatorship and democracy*. Cambridge University Press.

Akerlof, G. A. and R. E. Kranton (2005). Identity and the economics of organizations. *Journal of Economic Perspectives*, 9–32.

Alpizar, F., F. Carlsson, and O. Johansson-Stenman (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *Journal of Public Economics 92*(5-6), 1047–1060.

Andreoni, J. and R. Petrie (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics 88*(7-8), 1605–1623.

Bardhan, P. (2000). Irrigation and cooperation: An empirical analysis of 48 irrigation communities in South India. *Economic Development and Cultural Change 48*(4), 847–865.

Brosig, J. (2002). Identifying cooperative behavior: some experimental results in a prisoner's dilemma game. *Journal of Economic behavior and Organization 47*(3), 275–290.

Cardenas, J., J. Stranlund, and C. Willis (2000). Local environmental control and institutional crowding-out. *World Development 28*(10), 1719–1733.

Cooter, R. and B. Broughman (2005). Charity, Publicity, and the Donation Registry. *The Economists' Voice 2*(3), 4.

Fehr, E. and K. Schmidt (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the Economics of Giving, Altruism and Reciprocity 1*, 615–694.

Fehr, E., G. S. and G. Kirchsteiger (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica 65*, 833–860.

Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters 71*(3), 397–404.

Frank, R., T. Gilovich, and D. Regan (1993). The evolution of one-shot cooperation: An experiment. *Ethology & Sociobiology 14*(4), 247–256.

Frohlich, N. and J. Oppenheimer (1996). Experiencing impartiality to invoke fairness in the n-PD: Some experimental results. *Public Choice 86*(1), 117–135.

Glazer, A. and K. Konrad (1996). A signaling explanation for charity. *American Economic Review 86(4)*, 1019–1028.

Harbaugh, W. T. (1998, May). The prestige motive for making charitable transfers. *The American Economic Review 88*(2), 277–282.

Harrington Jr, J. (1989). If homo economicus could choose his own utility function, would he want one with a conscience? Comment. *The American Economic Review 79*(3), 588–593.

Harrison, G. (2002). Introduction to experimental economics. *At http://dmsweb. badm. sc. edu/glenn/manila/presentations*.

Hoge, D. R., C. E. Zech, M. J. Donahue, and P. H. McNamara (1996). *Money matters: Personal giving in American churches*. Westminster John Knox Pr.

Holmstrom, B. (1999). Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies 66*(1), 169–182.

Holt, C. and S. Laury (2009). Forthcoming. Theoretical explanations of treatment effects in voluntary contributions experiments. C. Plott, V. Smith, eds. *Handbook of Experimental Economic Results*.

Hugh-Jones, D. and D. Reinstein (2009). Secret santa: Anonymity in rituals.

Isaac James, M., R. Mark, and A. Williams (1994). Group size and the voluntary provision of public goods:: Experimental evidence utilizing large groups. *Journal of Public Economics 54*(1), 1–36.

Karlan, D. (2005). Using experimental economics to measure social capital and predict financial decisions. *American Economic Review 95*(5), 1688–1699.

Keser, C. and F. van Winden (2000). Conditional Cooperation and Voluntary Contributions to Public Goods. *Scandinavian Journal of Economics 102*(1), 23–39.

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (2001). Rational cooperation in the finitely repeated prisoners' dilemma. *Readings in Games and Information*.

Ledyard, J. (1993). *Public Goods: A Survey of Experimental Research*. Division of the Humanities and Social Sciences, California Institute of Technology.

Levy, G. (2007a). Decision making in committees: Transparency, reputation, and voting rules. *The American Economic Review 97*(1), 150–168.

Levy, G. (2007b). Decision-Making procedures for committees of careerist experts. *American Economic Review 97*(2), 306–310.

Londregan, J. and A. Vindigni (2006). Voting as a credible threat.

Martin, R., P. Fosh, H. Morris, P. Smith, and R. Undy (1991). The decollectivisation of trade unions? ballots and collective bargaining in the 1980s. *Industrial Relations Journal 22*(3), 197–208.

Milinski, M., D. Semmann, and H. Krambeck (2002). Reputation helps solve the 'tragedy of the commons'. *Nature 415*(6870), 424–6.

Orbell, J. M., A. J. C. V. D. Kragt, and R. M. Dawes (1988). Explaining discussion-induced cooperation. *Journal of personality and social psychology 54*(5), 811–819.

Ostrom, E. (2000a). Collective action and the evolution of social norms. *The Journal of Economic Perspectives*, 137–158.

Ostrom, E. (2000b). Collective action and the evolution of social norms. *The Journal of Economic Perspectives*, 137–158.

Plott, C. and V. Smith (2008). *Handbook of results in experimental economics*. North-Holland.

Prat, A. (2005). The wrong kind of transparency. *The American Economic Review 95*(3), 862–877.

Schram, A. (2000). Sorting out the seeking: The economics of individual motivations. *Public Choice 103*(3), 231–258.

Simpson, B. and R. Willer (2008). Altruism and Indirect Reciprocity: The Interaction of Person and Situation in Prosocial Behavior. *Social Psychology Quarterly 71*(1), 37–52.

Soetevent, A. (2005). Anonymity in Giving in a Natural Context: An Economic Field Experiment in Thirty Churches. *Journal of Public Economics 89*(11-12), 2301–2323.

Spence, M. (1973). Job Market Signaling. *Quarterly Journal of Economics 87*(3), 355–374.

Table 1: Summary statistics

	Stage 1 investment	Stage 2 invt.	Voted to exclude [a]	Min guess (Target: leader) [a]	Med. guess (Tgt: ldr) [a]	Max. guess (Tgt: ldr) [a]	Avg. guess (Tgt: follower) [a]	Votes against [b],[c]
Anonymous Treatment								
Mean	2.14	4.34	.377	2.64	4.36	5.79	4.32	.251
SD	1.21	3.24	.485	2.1	2.26	2.59	2.37	.222
P25	1	1	0	1	2.5	4	2	0
Median	2	5	0	2	4	6	5	.333
P75	3	7	1	4	6	8	6	.333
Min	0	0	0	0	0	0	0	0
Max	4	10	1	8	9	10	10	.667
Obs.	585	975	390	108	108	108	270	585
Revealed Treatment								
Mean	2.42	2.89	.495	2.21	3.49	4.67	3.46	.33
SD	1.02	2.7	.501	1.95	2.09	2.27	2.32	.61
P25	2	0	0	0	2	3	2	0
Median	2	2	0	2	3	5	3.5	0
P75	3	5	1	4	5	6	5	1
Min	0	0	0	0	0	0	0	0
Max	4	10	1	8	8	10	9	2
Obs.	585	975	390	108	108	108	270	585
Overall								
Mean	2.28	3.61	.436	2.43	3.93	5.23	3.89	.291
SD	1.13	3.07	.496	2.04	2.22	2.49	2.38	.46
P25	2	1	0	1	2	3.5	2	0
Median	2	3	0	2	4	5	4	0
P75	3	6	1	4	5	7	5.5	.333
Min	0	0	0	0	0	0	0	0
Max	4	10	1	8	9	10	10	2
Obs.	1170	1950	780	216	216	216	540	1170

[a] Observations for followers only, [b] ... for leaders only

[c] Anonymous case: set to $\frac{1}{3}$ the total votes (0,1,or 2) to exclude (for group/rep).

Table 2: Analysis of variance of stage 2 contributions by stage 1 contributions

Partial (marginal) sums of squares:						
	Anon.	Revealed	Anon. later	Rvld. later	Anon.	Rvld.
1 ecu invt.	14	14	35	7	16	21
2 ecu invt.	211	20	199	9.1	37	24
3 ecu invt.	605	42	428	15	36	28
4 ecu invt.	1008	88	497	19	75	44
Subject effects					2038	1560
Model Degrees of freedom	4	4	4	4	78	78
Observations	585	585	288	288	585	585
Model SS	1937	133	849	27	3975	1694
Total SS	6172	3997	2846	1519	6172	3997
R-sq.	.31	.033	.3	.018	.64	.42

'Later' refers to stages 8-15

Table 3: Probit regressions: decision to exclude someone

Dependent variable = Dummy: subject chooses to exclude someone.								
	(1)		(2)		(3)		(4)	
	All Repetitions		All Repetitions		Repetitions 8-15		Repetitions 8-15	
Minimum St. 1. Invt.	-0.075	(0.049)	-0.13*	(0.056)	-0.012	(0.065)	-0.10	(0.067)
Rvld × Min St. 1 Invt	-0.17*	(0.081)	-0.18*	(0.088)	-0.20+	(0.11)	-0.15	(0.11)
Median St. 1. Invt.	-0.011	(0.039)	-0.079+	(0.041)	-0.058	(0.053)	-0.11*	(0.054)
Rvld × Med. St. 1 Invt.	0.16*	(0.067)	0.19*	(0.080)	0.22*	(0.11)	0.21+	(0.12)
Range St 1. Invts.	-0.011	(0.041)	-0.040	(0.044)	0.028	(0.053)	-0.0027	(0.054)
Rvld × Range St. 1 Invts.	0.062	(0.062)	0.055	(0.071)	0.13	(0.092)	0.13	(0.093)
Revealed Treatment	-0.054	(0.17)	-0.018	(0.19)	-0.24	(0.22)	-0.072	(0.24)
History & Lag Var's	No		Yes		No		Yes	
Observations	780		660		384		384	

+ p<0.10, * p<0.05, ** p<0.01

Marginal effects at means reported. Std. errors (clustered by subject) in parentheses.

Table 4: Poisson regressions: exclusion votes against a player

Dependent variable = Number of votes [*] to exclude subject (in single repetition).						
	(1)		(2)		(3)	
	All Repetitions		Repetitions 8-15		All Repetitions	
St. 1 Investment	-0.13**	(0.049)	-0.17*	(0.071)		
Rvld × Invt.	-0.80**	(0.13)	-0.72**	(0.20)		
Repetition	0.015	(0.012)	-0.0047	(0.025)	0.014	(0.012)
Rvld × Reptn.	-0.012	(0.020)	0.035	(0.036)	-0.010	(0.022)
Invested 1 ecu					-0.24*	(0.12)
Rvld × Invt. 1 ecu					-0.29	(0.18)
Invested 2 ecu's					-0.32+	(0.17)
Rvld × Invt. 2 ecu's					-1.45**	(0.44)
Invested 3 ecu's					-0.56**	(0.20)
Rvld × Invt. 3 ecu's					-2.28**	(0.57)
Invested 4 ecu's					-0.52**	(0.17)
Rvld × Invt. 4 ecu's					-2.73**	(0.24)
Constant	-1.09**	(0.12)	-0.64+	(0.39)	-1.01**	(0.15)
Session/Trtmt. Dummies	Yes		Yes		Yes	
Observations	1170		576		1170	

+ p<0.10, * p<0.05, ** p<0.01

[*] Conditional expectation simulated for anonymous case; see footnote.

Poisson coef's: marginal effects. Robust (clustered by session/trtmt) SE's in parens.

Table 5: Poisson regressions: predictions for leaders

Dependent variable = Prediction of target's stage 2 investment						
	(1)	(2)	(3)	(4)	(5)	(6)
	All reps, Poisson	...	Reps 11,15	...	Rep 15	...
Target St. 1 Inv.	0.34**	0.25**	0.36**	0.24**	0.37**	0.27**
	(13.61)	(9.34)	(10.61)	(5.74)	(8.63)	(5.65)
Tgt. St. 1 Inv × Rvld.	-0.069+	-0.038	-0.091	-0.017	-0.22*	-0.15
	(-1.69)	(-0.97)	(-1.21)	(-0.21)	(-2.35)	(-1.55)
Repetition	-0.0077	0.0071				
	(-1.26)	(1.07)				
Rvld. × Repetition	-0.050**	-0.041**				
	(-4.65)	(-3.76)				
Dummy: revealed trtmt.	0.34*	0.33*	-0.25	-0.14	-0.044	0.0080
	(2.42)	(2.39)	(-1.12)	(-0.61)	(-0.17)	(0.03)
Constant	0.70**	0.19	0.55**	0.14	0.50**	0.079
	(7.87)	(1.58)	(5.63)	(1.23)	(4.21)	(0.48)
History & Lag 1 Var's	No	Yes	No	Yes	No	Yes
Observations	1152	1152	576	576	288	288
Sum: invt. & rvld.	.27**	.21**	.27**	.20**	.15+	.13+

+ p<0.10, * p<0.05, ** p<0.01

Session 1 excluded due to error in prediction instructions.

Poisson coef's: marginal effects. Robust s.e. (clustered by subject) in parens.

In anonymous treatments predictions were for (e.g.,) "the guy who contributed 4 ecus."

Table 6: Determinants of followers' stage 2 investment (Poisson regressions)

	(1)	(2)	(3)	(4)
	All reps	All reps	Reps 8-15	All reps
Min guess (target: leader)	0.11*			0.054
	(0.056)			(0.052)
... × Rvld. Trtmt.	0.065			0.080
	(0.087)			(0.062)
Med. guess (target: leader)	0.021			-0.029
	(0.068)			(0.060)
Max. guess (target: leader)	0.084			0.155**
	(0.061)			(0.044)
Guess (target: follower)	0.053			0.075**
	(0.038)			(0.022)
Num. Ldrs. Inv. 2+ [a]		0.205*	0.257+	0.104
		(0.092)	(0.139)	(0.180)
... × Rvld. Trtmt.		-0.081	-0.736*	-0.187
		(0.161)	(0.333)	(0.673)
Num. Ldrs. Inv. 3+ [b]		0.125+	0.149	-0.187+
		(0.066)	(0.103)	(0.112)
... × Rvld. Trtmt.		0.087	0.020	0.153
		(0.105)	(0.167)	(0.162)
Repetition	0.026	-0.035**	-0.036	-0.006
	(0.024)	(0.010)	(0.027)	(0.024)
Rvld × Reptn.	-0.030	-0.046**	-0.115**	0.013
	(0.029)	(0.015)	(0.038)	(0.025)
Final Rep.	-0.255	-0.115	-0.023	-0.157
	(0.210)	(0.170)	(0.171)	(0.246)
Dummy: Rvld. Trtmt.		-0.037	2.141**	-0.347
		(0.336)	(0.765)	(1.303)
Constant		1.184**	1.079*	0.119
		(0.196)	(0.492)	(0.392)
Subject-Fixed Effects	Yes [c]	No	No	No
Observations	125	780	384	192
P-val: F test, guesses	0.020			0.000
Sum coef:[d] Num. 3+ Ldrs., Anon		0.331**	0.407*	-0.083
Sum coef:[e] Num. 3+ Ldrs. Rvld		0.338*	-0.309	-0.116

+ p<0.10, * p<0.05, ** p<0.01. Standard errors in parentheses.

"Target" indicates the target of subject's prediction of round 2 gift for that repetition (3,7,11, or 13).

Pilot session dropped from regressions with subjects' prediction variables.

[a] Count of leaders (in group/rep.) who invested 2 or more in st. 1; excludes min. invt.

[b] ... 3 or more

[c] Subjects that were followers in 1 or fewer 'guessing' stages dropped from FE estimation; hence fewer obs.

[d] Net effect of adn'l leader investing 3+ in anon trtmt. Sums coef's for 'Num. Ldrs. Inv. 2+' and 'Num. Ldrs. Inv. 3+.'

[e] ... in revealed trtmt. 'Sum coef: Num. 3+ Ldrs' plus both '... x Rvld. Trtmt.' coef's

Figure 1: Experimental design

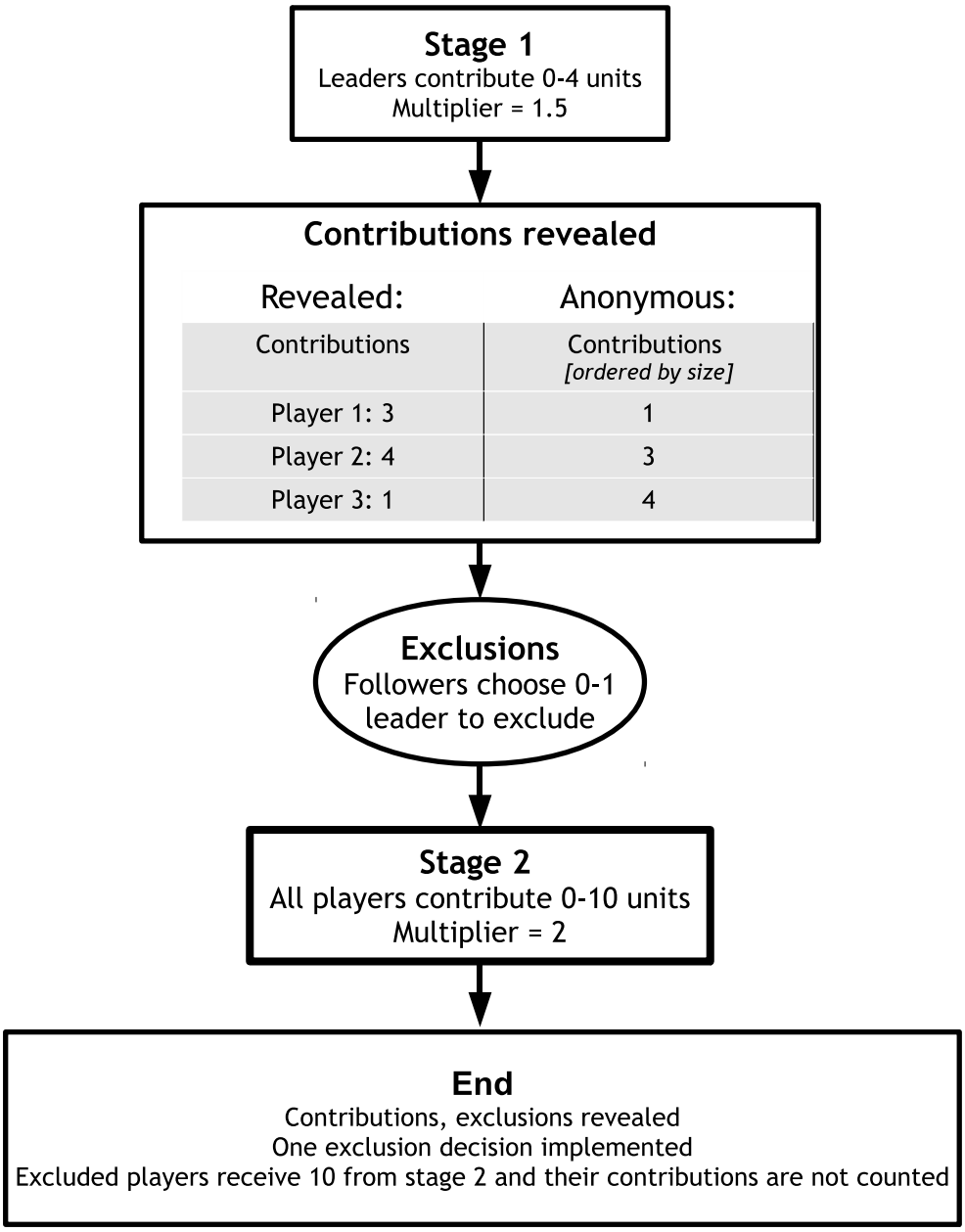


Figure 2: Mean contributions by repetition by treatment

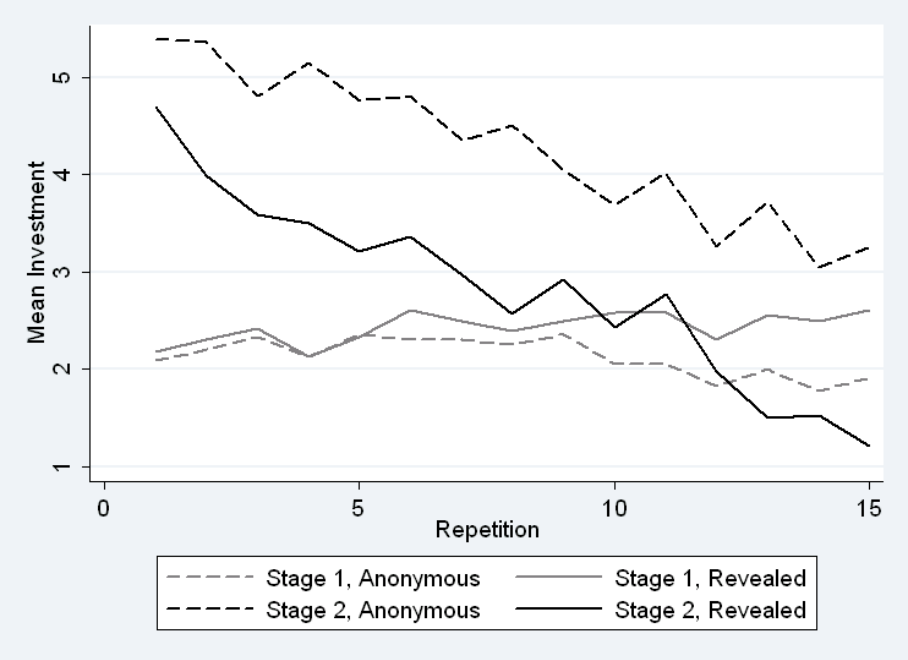


Figure 3: Mean contributions by session and treatment

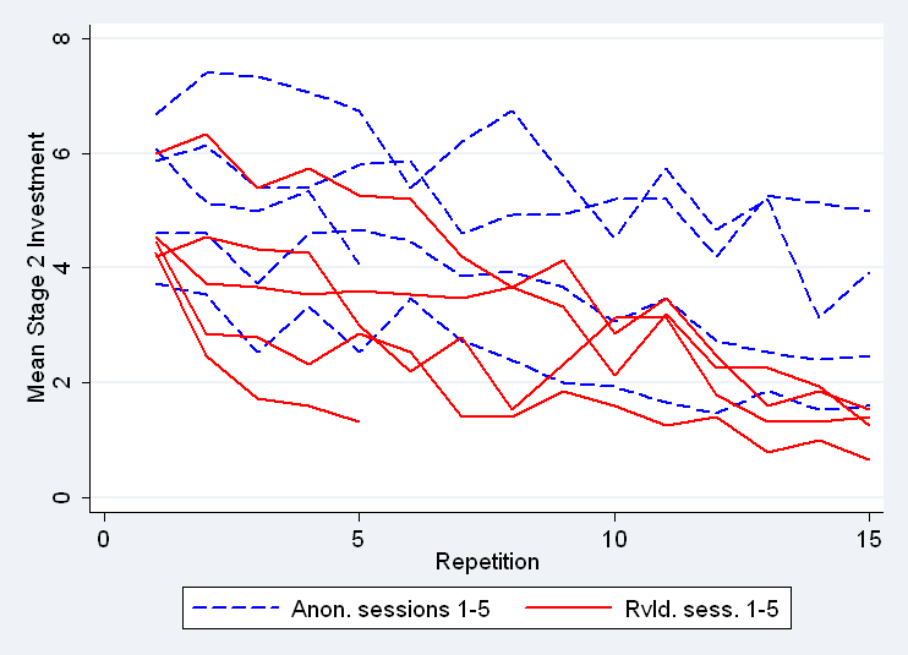


Figure 4: Stage 1 investment by stage 2 investment (repetitions 8-15). Left=anonymous, right=revealed

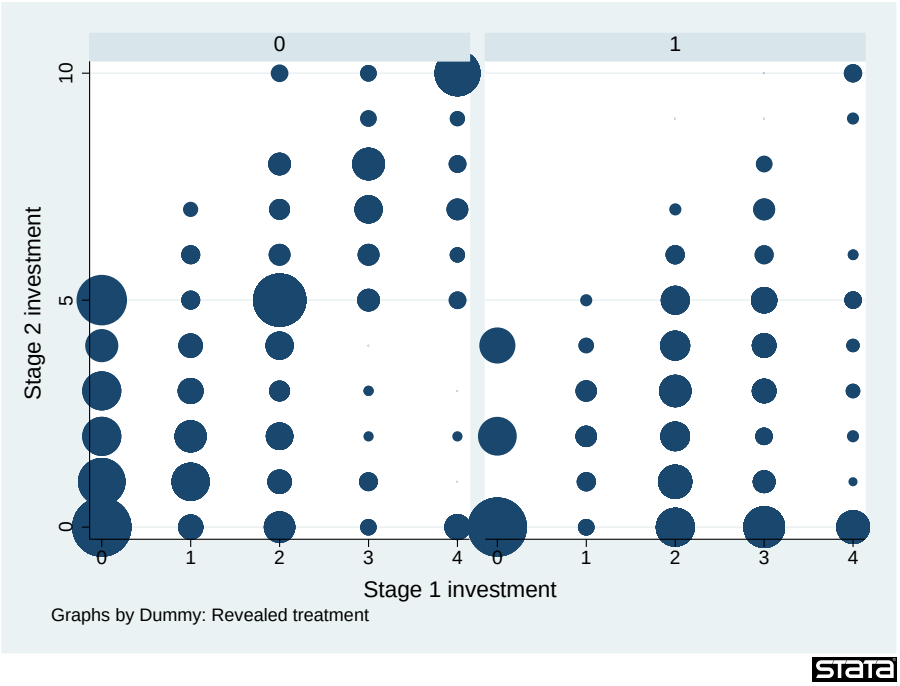


Figure 5: Predicted stage 2 contributions by (prediction target's) stage 1 contributions

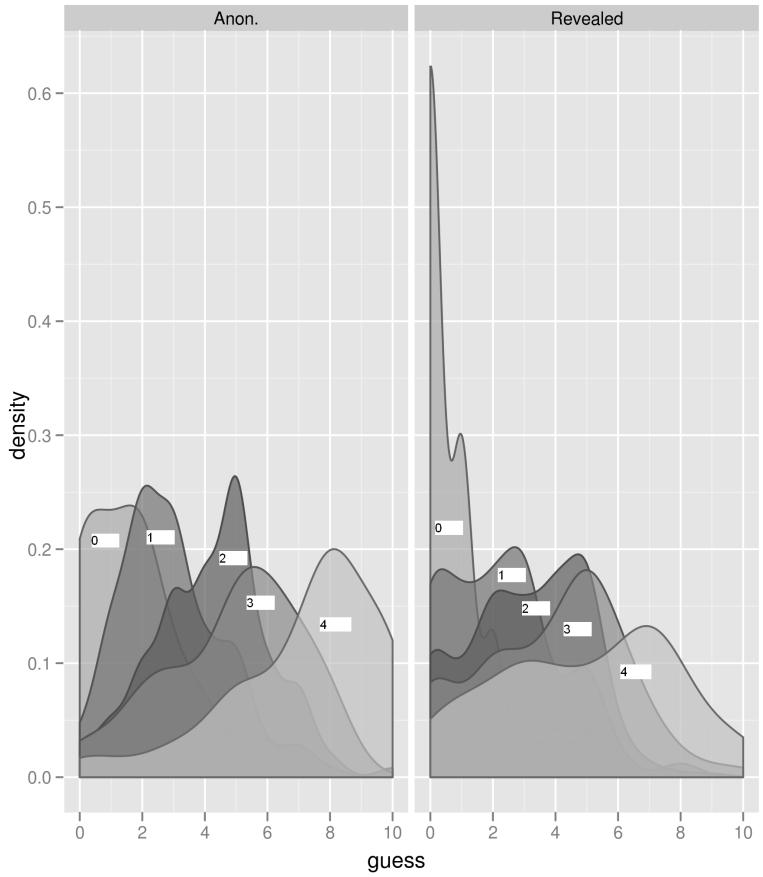
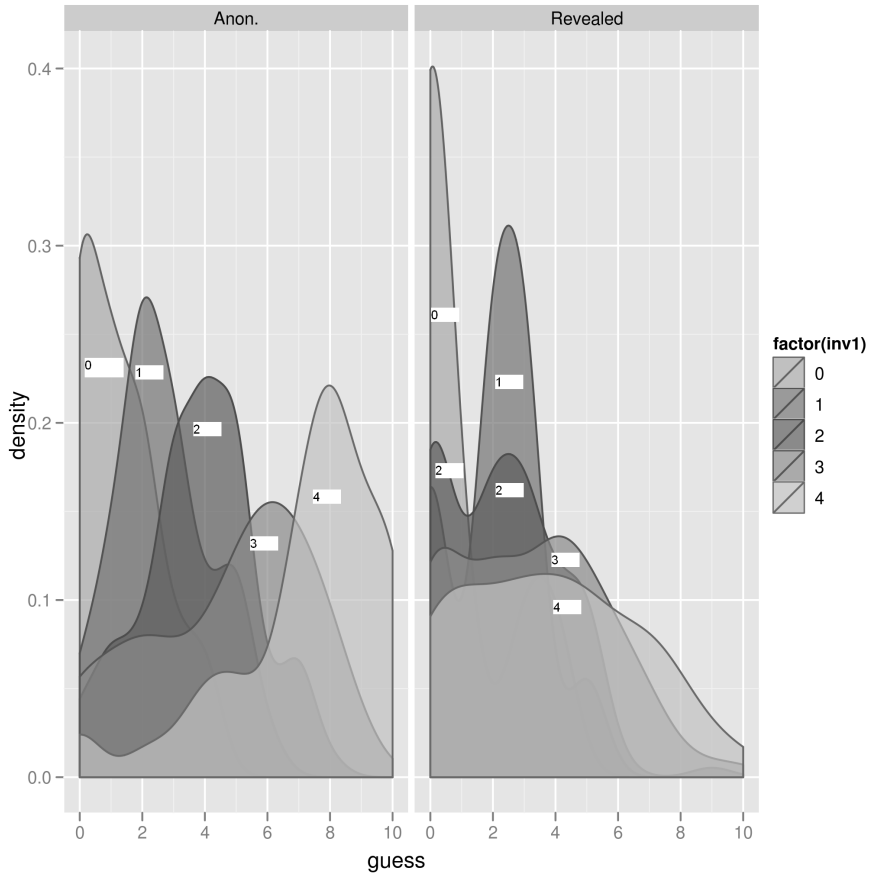


Figure 6: Predicted stage 2 contributions by (prediction target's) stage 1 contributions: repetitions 11 and 15



Appendix: Proofs

Notation

Define $\beta(x)$ as the best response to good-type donations of x when the distribution of good types is same as the prior. Hence, $\beta(x)$ satisfies

$$\sum \Pi(g)\psi_1(\beta(x), \frac{gx}{N-1}) = 1-\bar{\alpha}, \tag{6}$$

where $\Pi(g)$ is the probability of exactly g good types other than player i. Our conditions on ψ ensure that $\beta(0) > 0$ and $\beta'(x) > 0$ on $x \in [0, \bar{x}]$.

Proof that all equilibria are symmetric among good types:

Proof. When others' donations are uncertain, good types' optimal donations satisfy (4). Since ψ is strictly concave, this has a unique solution. Thus no good type plays a mixed strategy in equilibrium.

Suppose first that the number of good types is revealed to be $g+1$. We consider two cases: either the identity of the good types is known, or it is completely unknown. (These correspond to our revealed and anonymous signaling institutions.) Write π_{ij} for the probability, believed by any good type j, that player i is good. Players know their own type, so $\pi_{ii} = 1$. For other players, in the first case, $\pi_{ij} = 1$ for $i \in G$, a set of

g players, and 0 for all others; in the second case, $\pi_{ij} = \frac{g}{N-1}$ for all $i \neq j$. In either case, $\pi_{ij} = \pi_{ik}$ for $i \neq j$, $i \neq k$, $j \neq k$; pairs of good types share common probabilities about the type of third players. Also note that for $j \neq k$, $\pi_{jk} = \pi_{kj}$ if $j,k \in G$ in the identity known case, and always if identities are unknown.

Say that in equilibrium player i plays x_i if he is a good type. Let $j = \arg\max_i x_i$ and $k = \arg\min_i x_i$ (the maximum and minimum being taken over G in the identity known case). Suppose for a contradiction that $x_j > x_k$. Player j best responds to his expected distribution of others' donations, solving $E_{\bar{X}_{-j}} \psi_1(x_j, \bar{X}_{-j}) = 1 - \bar{\alpha}$, where $\bar{X}_{-j}$ is derived from the probabilities π_{ij} and donations x_i for all $i \neq j$. Similarly player k best responds to $\bar{X}_{-k}$. Now, since $\pi_{jk} = \pi_{kj}$, $x_j > x_k$ and all other probabilities and donations are common to both i and j, the distribution $\bar{X}_{-j}$ is first order stochastically dominated by $\bar{X}_{-k}$. But then, by $\psi_{12} > 0$ and $\psi_{11} < 0$, it must be that $x_j < x_k$, a contradiction.

The proof when the number of good types is unknown is similar and is omitted. □

Proof that x^* exists and is unique, and x_g exists and is unique for any g:

Proof. Suppose first that the number of good types is known to be $g+1$. Since all good types give the same in equilibrium, any point x_g is a fixed point of the continuous function $B(x_g) = b(\frac{g x_g}{N-1})$. By the Implicit Function Theorem applied to (3), $b' = -\frac{\psi_{12}}{\psi_{11}} > 0$. By our condition that $b'(X) = \frac{-\psi_{12}(x_g,X)}{\psi_{11}(x_g,X)} < k\frac{x_g}{X} = k\frac{N-1}{g}$, for $k < 1$, $B(\cdot)$ is a contraction on $[0,\bar{x}]$; also our conditions ensure that $B(x_g) \in [0,\bar{x}]$ for any x_g. Thus, B has a unique fixed point.

We define a symmetric equilibrium when good types are unknown, x^*, as a fixed point where $x^* = \beta(x^*)$. $x^* > 0$ since $\beta(0) > 0$, and it exists since $\beta(\bar{x}) \leq \bar{x}$ and β is continuous by the IFT. Implicitly differentiating (6) gives

$$\frac{d\beta(x)}{dx} = \frac{\sum \Pi(g) \frac{g}{N-1} \psi_{12}}{-\sum \Pi(g) \psi_{11}} > 0, \tag{7}$$

suppressing function arguments. By our condition on ψ_{12}/ψ_{11}, $\psi_{12}(\beta(x), \frac{g}{N-1}x) < -k(\beta(x)/\frac{g}{N-1}x)\psi_{11}$, so

$$\frac{\sum \Pi(g) \frac{g}{N-1} \psi_{12}}{-\sum \Pi(g) \psi_{11}} < \frac{\sum \Pi(g)(k\beta(x)/x)\psi_{11}}{\sum \Pi(g) \psi_{11}} \tag{8}$$

If $\beta(x) \leq x$, then $k\beta(x)/x < 1$ so the above is less than 1. Thus, if $\beta(x) \leq x$, then $\beta(x') < x'$ for $x' > x$. Therefore, $x^* = \beta(x^*)$ is unique.

□

Proof of Lemma 1:

Common knowledge of the number of good types increases donations ex ante when $\psi_1(x,X)$ is weakly concave in X and $b(\cdot)$ is weakly convex.

Proof. Define $\bar{g} = \sum \Pi(g)(g+1)$ as the expected total number of good types. First, suppose that $f(g) \equiv (g+1)x_g$ is convex. Then

$$\sum_{g=0}^{N-1} \Pi(g)(g+1)x_g \geq \sum_{g=0}^{N-1} \Pi(g)(g+1)x_{\bar{g}}; \tag{9}$$

and if $x^* < x_{\bar{g}}$[41], (5) follows immediately, i.e. knowledge increases contributions.

We next prove that if the Lemma conditions hold, both the above conditions hold: $f(g)$ is convex and $x^* < x_{\bar{g}}$.

To show $x^* < x_{\bar{g}}$, first observe that

$$\psi_1(x^*, \frac{\bar{g}x^*}{N-1}) \geq E_g \psi_1(x^*, \frac{gx^*}{N-1}) = 1-\bar{\alpha}, \tag{10}$$

the inequality by concavity of ψ_1, the equality by definition of x^*. Now suppose $x^* \geq x_{\bar{g}}$. Then $\psi_1(x^*, \frac{\bar{g}x^*}{N-1}) < 1-\bar{\alpha}$, a contradiction. (Proof: $b'(\frac{\bar{g}x_{\bar{g}}}{N-1}) < \frac{N-1}{\bar{g}}$, as in the previous proof. So for some small ε and all $x \in (x_{\bar{g}}, x_{\bar{g}}+\varepsilon)$, $b(\frac{\bar{g}x}{N-1}) < x$. Suppose $b(\frac{\bar{g}x^*}{N-1}) \geq x^*$. Then at some point $y \in (x_{\bar{g}}, x^*]$, $b(\frac{\bar{g}y}{N-1}) = y$. But this would contradict uniqueness of $x_{\bar{g}}$. So $b(\frac{\bar{g}x^*}{N-1}) < x^*$. Then, since $\psi_{11} < 0$, we have $\psi_1(x^*, \frac{\bar{g}x^*}{N-1}) < \psi_1(b(\frac{\bar{g}x^*}{N-1}), \frac{\bar{g}x^*}{N-1}) = 1-\alpha$.) Thus $x^* < x_{\bar{g}}$.

To show $f(g) = (g+1)x_g$ is convex, it suffices to show that x_g is convex. Now x_g solves $b(\frac{gx_g}{N-1}) - x_g = 0$. Applying the Implicit Function Theorem,

$$\frac{dx_g}{dg} = \frac{-\frac{x_g}{N-1}b'(\frac{gx_g}{N-1})}{\frac{g}{N-1}b'(\frac{gx_g}{N-1}) - 1}. \tag{11}$$

This is positive, by $b' > 0$ and $b' < \frac{N-1}{g}$, and we can rearrange it to

$$\frac{dx_g}{dg} = \frac{x_g/(N-1)}{\frac{1}{b'(gx_g/(N-1))} - \frac{g}{N-1}} > 0. \tag{12}$$

Now if g increases, then the top increases while the denominator decreases, since b' is weakly increasing by convexity of b. Thus dx_g/dg increases in g, showing that x_g is convex. □

The following Lemma is required for our proposition.

Lemma. *In the minigame, a separating equilibrium in which good types donate $x < \beta(x)$ in the first round is not intuitive.*

Proof. In any separating equilibrium, bad types donate less than good types, since otherwise they could increase their first round utility and simultaneously pool with good types, inducing greater contributions in the later round. Thus, bad types donate $y < x$. Now say $x < \beta(x)$ and consider a deviation by a good type to $\beta(x)$. Since bad types prefer donating y and being recognized as a bad type to donating x and being recognized as a good type for sure, a fortiori they would not prefer to donate $\beta(x) > x$ whatever the resulting belief. Good types, however, would prefer to donate $\beta(x)$ than x, since $\beta(x)$ is a the good type's best response when other good types are donating x and the distribution of good types is the prior. In particular, if the resulting belief is that the player is good for sure (or, in the anonymous case, that there is one more good type), then good types prefer to deviate to $\beta(x)$. But if so, good types have a credible deviation and the equilibrium is not intuitive. □

[41] The definition of x_g can be extended unchanged to non-integer values of g.

Proposition. *In the anonymous minigame, there is an Intuitive separating equilibrium if and only if $D \geq \hat{D}$. In the revealed minigame, there is an Intuitive separating equilibrium if and only if $D \geq D^*$, where $D^* > \hat{D}$.*

Proof. By the previous Lemma we can assume that good types donate $x \geq \beta(x)$. In the revealed game, suppose there is a separating equilibrium where good types donate $x \geq \beta(x)$ in the first round, bad types donate 0. Beliefs are such that those who donate x are believed good with 100% probability; those donating less than x are believed bad with 100% probability; beliefs can be anything for those donating more than x. These beliefs support play as specified, *if* there is separation in equilibrium: good types cannot do better than playing x, since $x \geq \beta(x)$ and their utility is concave, and bad types cannot do better than playing 0.

Good type donations, after g good types are revealed and only these players are included, are y_g satisfying $\psi_1(y_g, y_g) = 1 - \alpha/g$. Therefore, the bad type's incentive compatibility constraint (IC) to play 0 instead of x is

$$\sum_{g=0}^{N-1} \Pi(g) D\alpha \frac{g}{N} x \geq \sum_{g=0}^{N-1} \Pi(g) \left\{ D(\alpha \frac{g+1}{N} - 1)x + \alpha \frac{g}{g+1} y_{g+1} \right\}. \tag{13}$$

In the second round, if the bad type gives x, he will be included in a group of $g+1$, of whom g will give y_{g+1}. Simplifying this:

$$D(1 - \frac{\alpha}{N})x \geq \sum_{g=0}^{N-1} \Pi(g) \alpha \frac{g}{g+1} y_{g+1}. \tag{14}$$

For any x equilibrium we can now calculate the lowest D that satisfies the bad type IC. This will make the above hold with equality.

We show in 3 that when (14) is satisfied, the good type's IC is also satisfied. Therefore, the lowest D allowing for separation in the revealed institution will satisfy (14) with equality. The lowest D possible is when $x = 1$, giving:

$$D^* = \frac{\sum_{g=0}^{N-1} \Pi(g) \alpha \frac{g}{g+1} y_{g+1}}{1 - \alpha/N}. \tag{15}$$

Next we show that these separating equilibria are intuitive. For any $x \in [\beta(x), 1]$, if (14) holds with equality, the bad type is just indifferent between playing 0 and playing x. Thus, he would strictly prefer to play $y \in (0, x)$ if this would result in him being believed good for sure. Therefore, the good type has no credible deviation to $y < x$. The good type could credibly deviate to $y > x$ (since bad types would not do this for any resulting belief) but has no incentive to: since $x > \beta(x)$ and good type utility is concave, deviating to $y > x$ would reduce round 1 utility and could not improve on the belief the good type induces by playing x. So far we have ensured that an equilibrium with $x \in [\beta(x), 1]$ and D such that (14) holds with equality is indeed intuitive. For even higher values of D, there is an equilibrium with $x = \beta(x) = x^*$ and (14) holding with strict inequality. Then good types have no incentive to deviate to any $y \neq x$, and bad types have no incentive to deviate to any $y > 0$. Thus, there is always an intuitive separating equilibrium for $D \geq D^*$.

Now, we turn to the anonymous institution and again seek conditions for a separating equilibrium. Thus, after g players are revealed as good types, all N players are included and good type donations are x_{g-1}. The bad type IC is

$$\sum_{g=0}^{N-1} \Pi(g) \left\{ D \frac{g}{N} \alpha x + \alpha \frac{g}{N} x_{g-1} \right\} \quad \geq \quad \sum_{g=0}^{N-1} \Pi(g) \left\{ D(\frac{g+1}{N} \alpha - 1)x + \alpha \frac{g}{N} x_g \right\}. \tag{16}$$

Rearranging, this becomes

$$D\left(1-\frac{\alpha}{N}\right)x \geq \sum_{g=0}^{N-1} \Pi(g)\alpha\frac{g}{N}\left(x_g - x_{g-1}\right). \tag{17}$$

To show that for any x, the lowest D satisfying this will be less than the lowest D satisfying (14), it will suffice to show that $x_g \leq y_{g+1}$ for all g. Since $\psi_1(x_g, \frac{g}{N-1}x_g) = 1-\alpha/N$, by the positive cross-partial we have $\psi_1(x_g, x_g) > 1-\alpha/N \geq 1-\alpha/(g+1) = \psi_1(y_{g+1}, y_{g+1})$. But then $x_g < y_{g+1}$. (Proof: by assumption, $\frac{\psi_{12}(x,x)}{\psi_{11}(x,x)} > -\frac{x}{x} = -1$, so that $\frac{d}{dx}\left(\psi_1(x,x)\right) = \psi_{12}(x,x) + \psi_{11}(x,x) < 0$.)

It remains only to prove that when the bad type's IC (17) is satisfied with equality, the good type IC is satisfied. This is shown in Lemma 4 in the Online Appendix.

The arguments that the separating equilibria in the anonymous institution are intuitive closely parallel those for the revealed institution, and are omitted. □

Online Appendix: Supplemental results and proofs

Figure 7: Mean net profits by period and treatment

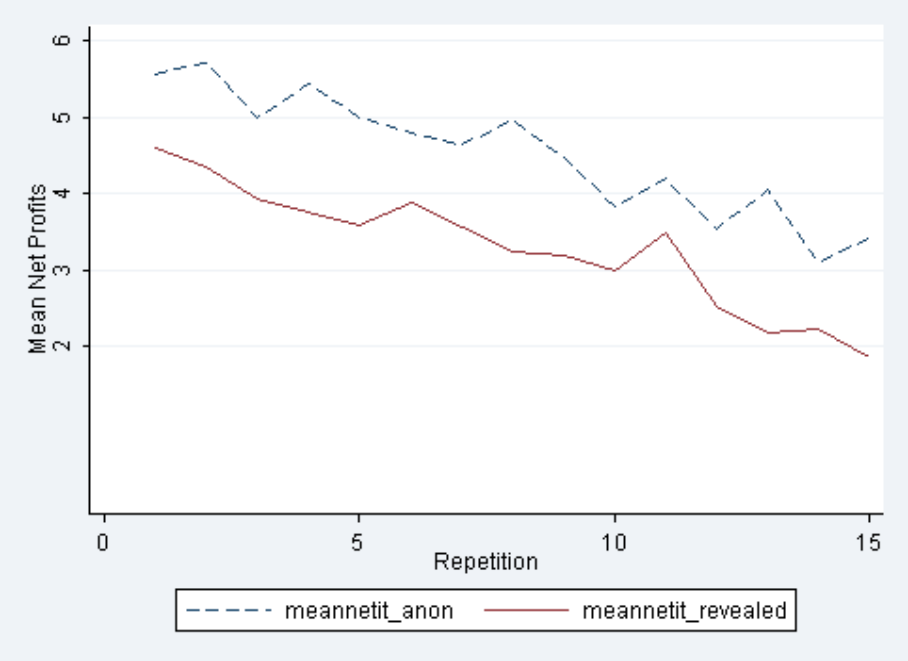


Figure 8: Predictions by own contributions

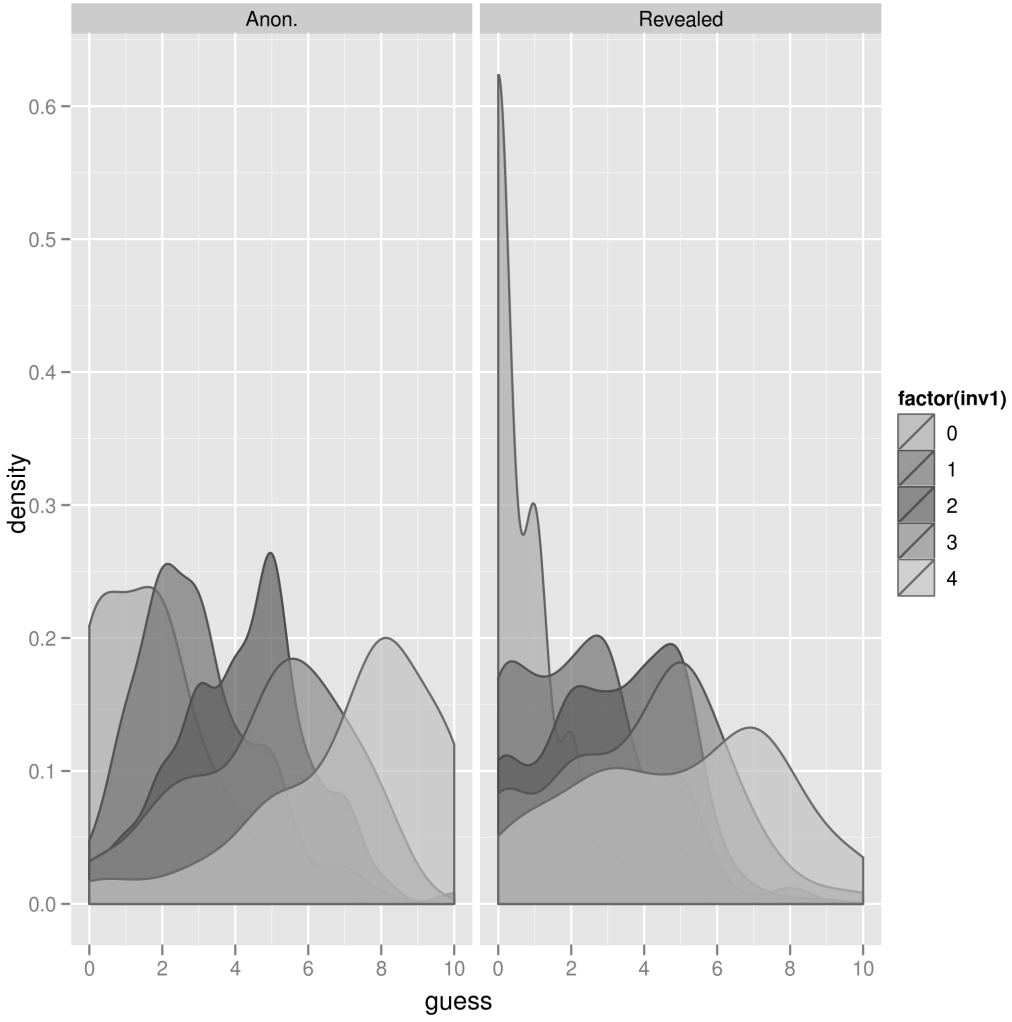


Table 7: Regressions: Stage 2 contributions by stage 1 contributions

	(1) OLS- Anon.	(2) OLS-Revealed	(3) OLS- Anon. later	(4) OLS-Revealed later	(5) OLS-FE Anon.	(6) OLS-FE Revealed
Invested 1 ecu	0.62	0.87	1.28**	1.08	0.82+	1.27*
	(0.50)	(0.66)	(0.39)	(0.66)	(0.43)	(0.60)
Invested 2 ecus	2.25**	0.92	2.92**	1.10	1.18**	1.35*
	(0.53)	(0.69)	(0.47)	(0.70)	(0.40)	(0.59)
Invested 3 ecus	4.19**	1.34+	4.85**	1.42+	1.30**	1.57*
	(0.63)	(0.73)	(0.62)	(0.73)	(0.45)	(0.63)
Invested 4 ecus	5.33**	2.06*	5.14**	1.69+	2.09**	2.08**
	(0.76)	(0.89)	(0.97)	(0.94)	(0.50)	(0.67)
Constant	1.91**	1.63*	0.97**	0.75	3.31**	1.35*
	(0.46)	(0.65)	(0.29)	(0.63)	(0.35)	(0.56)
Observations	585	585	288	288	585	585

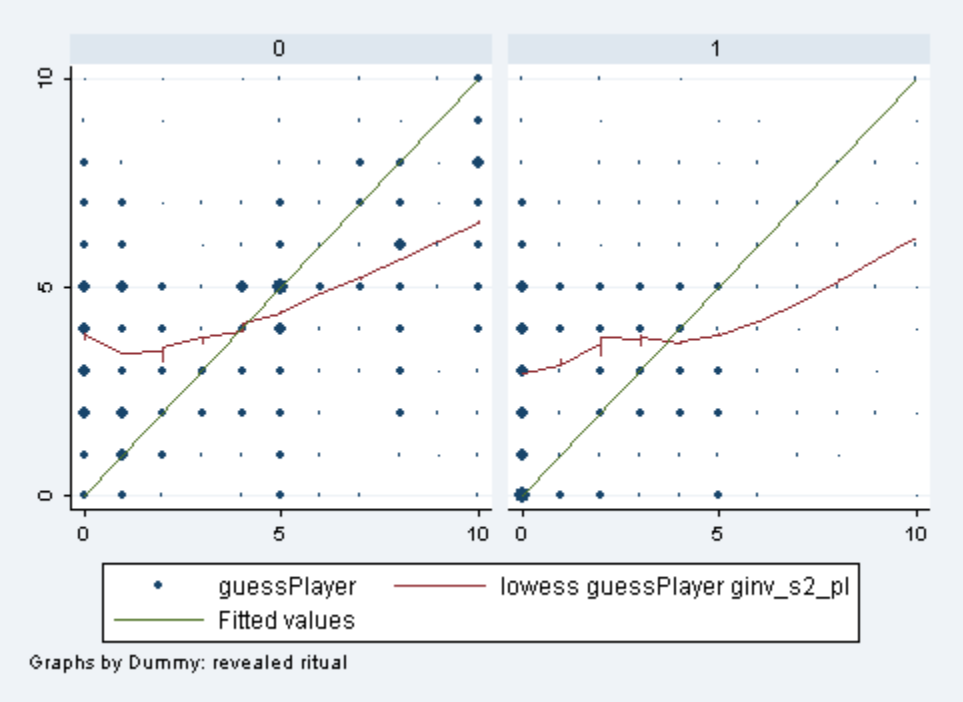
+ p<0.10, * p<0.05, ** p<0.01

Robust standard errors (clustered by subject) in parentheses.

Later refers to stages 8-15

Notes on table 7: Note that the coefficient on stage 1 investment is significant and positive in the anonymous case, but significantly smaller, and sometimes insignificant, in the revealed case. Again, a 1 ecu investment explains little, while the coefficients on the larger anonymous investment dummies are significant, suggesting a nonlinear relationship.

Figure 9: Predicted gift by actual (partner's) gift, Left=anonymous, right=revealed



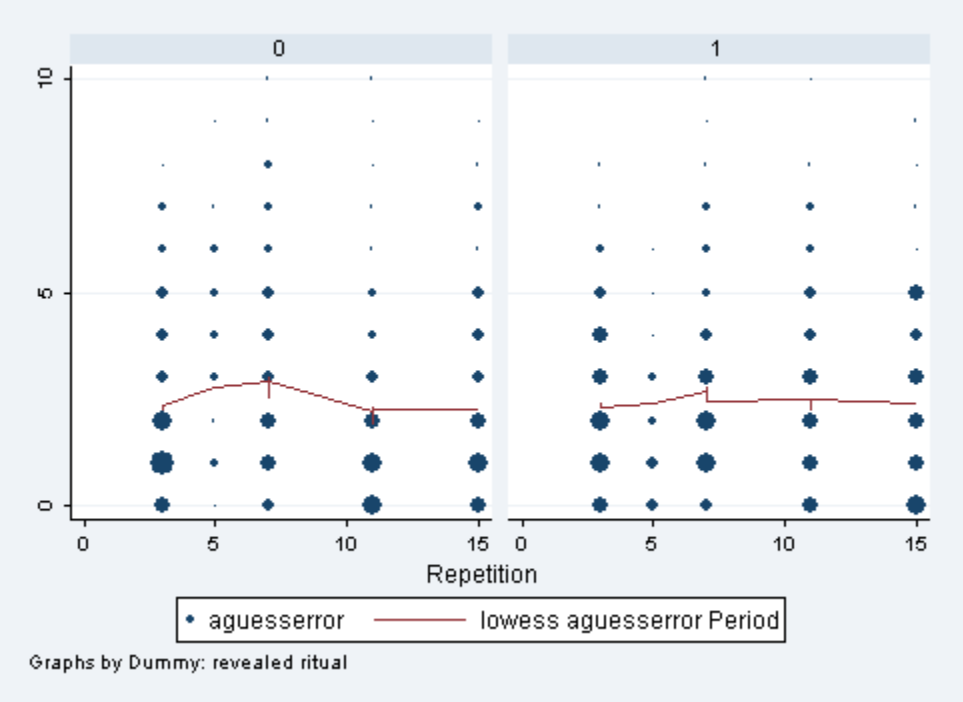
Horizontal axis represents partner's actual gift, vertical axis represents prediction for partner.

Width of bubbles increase in number of observations at point.

Lines fitted with lowess smoothing.

Added: 45 degree line of "perfect prediction".

Figure 10: Error in prediction (absolute value) by repetition



Horizontal axis represents repetition (3,5,7,11, or 15),
vertical axis represents absolute value of subject's prediction error (for partner).
Width of bubbles increase in number of observations at point.
Lines fitted with lowess smoothing.

Table 8: Subjects' prediction accuracy and bias

Treatment	*Anon.*	*Revealed*	*Difference*$^{(A)}$	*Anon*	*Revealed*	*Dfc.*$^{(A)}$
Repetitions	*3,5,7,11,15*	*3,5,7,11,15*	*3,5,7,11,15*	*11,15*	*11,15*	*11,15*
Mean error$^{(B)}$	0.14	0.86**	-0.72**,**	0.03	0.87**	-0.83**,**
(std. error)	(.13)	(.12)	(.18)	(.17)	(.18)	(.25)
Mean Absolute error	2.43**	2.43**	0.00	2.08**	2.44**	-0.36*,*
(std. error)	(.08)	(.08)	(.12)	(.12)	(.13)	(.17)
Mean Normalized error$^{(C)}$	1.08**	1.64**	-0.55**	0.84**	2.43**	-1.59**,**
(std. error)	(.01)	(.03)	(.03)	(.00)	(.05)	(.04)
Corr(predicted, actual)	0.39**	0.26**	[.19]$^{(D)}$	0.52**	0.10	[.00**]$^{(D)}$

+ p<0.10, * p<0.05, ** p<0.01, standard 2-tailed significance tests (t-tests).

(After comma) + p<0.10, * p<0.05, ** p<0.01, significance in 2-tailed rank-sum tests.

(A) "Difference" is for previous two columns, i.e., value for anonymous minus value for revealed treatment.

(B) "Error" is subject's prediction for partner minus partner's actual contribution.

(C) "Normalized error" is mean of squared error in subject's guess divided by the variance of this.

(D) Brackets: P-value, significance of interaction term from corresponding regression, robust standard error (clustered by id).

Notes on table 8: Subjects were overoptimistic in both treatments, significantly so only in the revealed treatment, in which they over-predicted by an average of 0.86 ecus; predictions did not strongly improve in later stages (see also Figures 9 and 10 and Table 8, online appendix).

Table 9: Poisson regressions: controlling for previous exclusion

Dependent variable: Subject's stage 2 investment. Repetitions 8-15, followers only.			
	(1)	(2)	(3)
	Poisson	Poisson fe	Poisson
Revealed	2.01**		1.82*
	(0.70)		(0.75)
Count: Leaders who gave 2+	0.25+	0.13	0.19
	(0.13)	(0.10)	(0.13)
Revealed×...	-0.74*	-0.74*	-0.65*
	(0.31)	(0.35)	(0.33)
Count: Leaders who gave 3+	0.10	0.03	-0.03
	(0.10)	(0.076)	(0.10)
Revealed×...	0.080	-0.14	0.11
	(0.17)	(0.13)	(0.16)
Repetition	-0.025	-0.04+	-0.01
	(0.03)	(0.02)	(0.03)
Rvld. × Repetition	-0.14**	-0.14**	-0.12**
	(0.04)	(0.03)	(0.04)
Was excluded	-0.35+	-0.07	-0.16
	(0.19)	(0.20)	(0.24)
Revealed×...	0.57*	-0.01	0.19
	(0.28)	(0.55)	(0.35)
Times excluded			-0.11
			(0.12)
Revealed×...			0.20
			(0.14)
Lag others' ctrbn controls	No	No	Yes
Constant	1.20**		0.62
Observations	384	315	384
Sum coef:Num. 3+ Ldrs., Anon	0.35*	0.16	0.16
Sum coef: Num. 3+ Ldrs. Rvld	-0.31	-0.72*	-0.38+
Sum coef: Rvld. × excluded	0.22	-0.08	0.03

+ p<0.10, * p<0.05, ** p<0.01

Std. err. in parens (clustered by session/treatment for non-FE models).

Notes on table 9: To test the "embitterment" hypothesis we run three Poisson regressions of second stage investment for follower subjects in repetitions 8-15. We examine the latter stages as behavior is more likely to have converged, and where there is enough experience for embitterment to be a possibility. We focus on follower subjects to be consistent with table 6 – however, the findings below are similar for leader subjects (available by request). The first column is a standard Poisson regression, the second column includes a subject-fixed effect, and the final column controls for three lags of "other subjects' second round contributions" In each of these we include a dummy variable "was excluded" indicating whether a subject has been excluded in any previous repetition" the final column also controls for the *number* of times a subject was previously excluded.. The interaction of these dummies with the revealed treatment is given by "Revealed× ". The net coefficient on previous exclusion for the revealed treatment ("Sum coef: Rvld. × excluded") is positive in the first and third columns and insignificant in all columns – this offers evidence against an embitterment effect. Even with these controls the stage-2 contributions decline significantly more rapidly in the revealed treatment ("Rvld. × Repetition" coefficients). Finally, the net effect of an additional leader giving three or more remains positive in the anonymous case ("Sum coef: Num. 3+ Ldrs., Anon") although it is only significant in the first column.

IV regression results : see file xtivregresults.doc

Screenshots

See file: screenshots.zip. (Translation by request).

Proofs

Lemma 3. *In the revealed institution, the good type's incentive compatibility condition holds when the bad type's IC condition* (14) *holds with equality.*

Proof. The good type's IC is

$$\sum_{g=0}^{N-1}\Pi(g)D\left[\alpha\frac{gx+\beta(x)}{N}-\beta(x)+\psi(\beta(x),\frac{g}{N-1}x)\right]\leq \tag{18}$$
$$\sum_{g=0}^{N-1}\Pi(g)\left\{D\left[\alpha\left(\frac{g+1}{N}-1\right)x+\psi(x,\frac{g}{N-1}x)\right]+[\alpha y_{g+1}-y_{g+1}+\psi(y_{g+1},y_{g+1})]\right\}.$$

Here, the left hand side is the benefit from playing the first round best response $\beta(x)$ rather than x, and thus being excluded in the second round. The right hand side is the benefit from playing x and being included in the second round with g other good types, whereupon everyone plays y_{g+1}. Simplifying this gives

$$\sum_{g=0}^{N-1}\Pi(g)D\left[(\frac{\alpha}{N}-1)(\beta(x)-x)+\psi(\beta(x),\frac{g}{N-1}x)-\psi(x,\frac{g}{N-1}x)\right]\leq \tag{19}$$
$$\sum_{g=0}^{N-1}\Pi(g)\left\{(\alpha-1)y_{g+1}+\psi(y_{g+1},y_{g+1})\right\}.$$

Will this be satisfied when the bad type IC just holds? Since $\psi_1>1$ and $x\geq\beta(x)$ the left hand side is less than $\sum_{g=0}^{N-1}\Pi(g)D(\frac{\alpha}{N}-1)(\beta(x)-x)\equiv D(1-\frac{\alpha}{N})(x-\beta(x))$. So the above will be satisfied if

$$D(1-\frac{\alpha}{N})(x-\beta(x))\leq\sum_{g=0}^{N-1}\Pi(g)\left\{(\alpha-1)y_{g+1}+\psi(y_{g+1},y_{g+1})\right\} \tag{20}$$

equivalently

$$D(1-\frac{\alpha}{N})x\leq D(1-\frac{\alpha}{N})\beta(x)+\sum_{g=0}^{N-1}\Pi(g)\left\{(\alpha-1)y_{g+1}+\psi(y_{g+1},y_{g+1})\right\}. \tag{21}$$

When the bad type IC just holds, we can replace the left hand side using (14), to give

$$\sum_{g=0}^{N-1}\Pi(g)\alpha\frac{g}{g+1}y_{g+1}\leq D(1-\frac{\alpha}{N})\beta(x)+\sum_{g=0}^{N-1}\Pi(g)\left\{(\alpha-1)y_{g+1}+\psi(y_{g+1},y_{g+1})\right\}. \tag{22}$$

Now,

$$\psi(y_{g+1}, y_{g+1}) = \psi(0, y_{g+1}) + \int_0^{y_{g+1}} \psi_1(y, y_{g+1})dy > [1-\alpha/(g+1)]y_{g+1}, \tag{23}$$

by the FOC on y_{g+1} and concavity of ψ. So the right hand side is greater than

$$\sum_{g=0}^{N-1} \Pi(g) \{(\alpha-1)y_{g+1} + [1-\alpha/(g+1)]y_{g+1}\} = \sum_{g=0}^{N-1} \Pi(g)\alpha\frac{g}{g+1}y_{g+1} \tag{24}$$

and thus (22) holds with strict inequality. □

Lemma 4. *In the anonymous institution, the good type's incentive compatibility condition holds when the bad type's IC condition* (17) *holds with equality.*

Proof. The good type IC is

$$\begin{aligned} &\sum_{g=0}^{N-1} \Pi(g) \left\{ D\left[\alpha\tfrac{g+1}{N}x - x + \psi(x, \tfrac{g}{N-1}x)\right] + \alpha\tfrac{g+1}{N}x_g - x_g + \psi(x_g, \tfrac{g}{N-1}x_g)\right\} \geq \\ &\sum_{g=0}^{N-1} \Pi(g) \left\{ D\left[\alpha\left(\tfrac{g}{N}x + \tfrac{\beta(x)}{N}\right) - \beta(x) + \psi(\beta(x), \tfrac{g}{N-1}x)\right] + \alpha\left(\tfrac{g}{N}x_{g-1} + \hat{z}\right) - \hat{z} + \psi(\hat{z}, \tfrac{g}{N-1}x_{g-1})\right\} \end{aligned} \tag{25}$$

where $\hat{z}$ is a best response to g other good types who each donate x_{g-1}; $\hat{z} = b(\frac{g}{N-1}x_{g-1}) \in (b(\frac{g-1}{N-1}x_{g-1}), b(\frac{g}{N-1}x_g)) \equiv (x_{g-1}, x_g)$ by increasingness of $b(\cdot)$. Rearranging, this becomes:

$$\begin{aligned} &D\left[\left(1-\frac{\alpha}{N}\right)(x-\beta(x)) + \sum_{g=0}^{N-1}\Pi(g)\left\{\psi(\beta(x), \frac{g}{N-1}x) - \psi(x, \frac{g}{N-1}x)\right\}\right] \leq \\ &\sum_{g=0}^{N-1}\Pi(g)\left\{\alpha\frac{g}{N}(x_g - x_{g-1}) - \left(1-\frac{\alpha}{N}\right)(x_g - \hat{z}) + \psi(x_g, \frac{g}{N-1}x_g) - \psi(\hat{z}, \frac{g}{N-1}x_{g-1})\right\} \end{aligned} \tag{26}$$

and using $\psi(\beta(x), \frac{g}{N-1}x) - \psi(x, \frac{g}{N-1}x) \leq 0$ by $\psi_1 > 0$ and $\beta(x) \leq x$, the left hand side is less than

$$D\left(1-\frac{\alpha}{N}\right)(x-\beta(x)) < D\left(1-\frac{\alpha}{N}\right)x. \tag{27}$$

On the right hand side, we can write

$$\begin{aligned} & \textstyle\sum_{g=0}^{N-1}\Pi(g)\left[\psi(x_g, \frac{g}{N-1}x_g) - \psi(\hat{z}, \frac{g}{N-1}x_{g-1})\right] \\ = \; & \textstyle\sum_{g=0}^{N-1}\Pi(g)\left[\psi(x_g, \frac{g}{N-1}x_g) - \psi(\hat{z}, \frac{g}{N-1}x_g) + \psi(\hat{z}, \frac{g}{N-1}x_g) - \psi(\hat{z}, \frac{g}{N-1}x_{g-1})\right] \\ = \; & \textstyle\sum_{g=0}^{N-1}\Pi(g)\left[\int_{\hat{z}}^{x_g}\psi_1(\bar{z}, \frac{g}{N-1}x_g)d\bar{z} + \psi(\hat{z}, \frac{g}{N-1}x_g) - \psi(\hat{z}, \frac{g}{N-1}x_{g-1})\right] \\ > \; & (x_g - \hat{z})\left(1-\tfrac{\alpha}{N}\right) \end{aligned} \tag{28}$$

by the FOC for x_g , concavity of ψ and $\psi_2 > 0$. Plugging this inequality into (26) shows that the right hand side is greater than

$$\sum_{g=0}^{N-1}\Pi(g)\alpha\frac{g}{N}(x_g - x_{g-1}). \tag{29}$$

Putting this together with the bound (27) on the LHS of (26) , we find that (26) holds with strict inequality

so long as

$$D\left(1-\frac{\alpha}{N}\right)x \leq \sum_{g=0}^{N-1} \Pi(g)\alpha\frac{g}{N}\left(x_g - x_{g-1}\right) \tag{30}$$

which holds when the bad type IC condition (17) holds with equality. □