Do I care what you think of me? Varying observability in a public goods experiment

Hannes Titeca*
University of Exeter

December 2, 2020

Abstract

In social dilemmas the level of observability by others often differs, either by design or by coincidence, and it seems intuitive that this might affect prosocial behaviour. For example; the potential for negative emotions such as guilt or shame may be greater when behaviour is more easily observed by others. A 3×1 between–subject design is used where the level of ex-post disclosure/feedback others receive is varied in a lab–based, anonymous, one–shot, two–player public goods game. Incentivised first order beliefs (what a subject believes their partner will contribute) and second order beliefs (what a subject believes their partner expects them to contribute) are directly elicited. A within–subjects analysis, mitigating any consensus effect concerns (people believing that others think and act like them), finds that at least some people are "shame averse"; when able to be observed ex–post, there is a stronger response to their second order belief of what they believe their partner expects. The same analysis also finds robust evidence for guilt aversion which is possible even when there is no observability by others. When available, receiving information on the average past behaviour of others has the effect of reinforcing any shift in the observed average behaviour.

Keywords: shame aversion, disclosure, observation, guilt aversion, public goods games, psychological game theory

*This work was supported by the Economic and Social Research Council. I would also like to thank, in no particular order, the University of Exeter, my PhD colleagues, those at various events with whom I discussed this work, and, lastly, my supervisors Brit Grosskopf, Sebastian Kripfganz and Rajiv Sarin for their continued support and advice.

1 Introduction

This paper, along with much of the work it builds upon, assumes that at least some people have belief–dependent motivations. An example of this is behaviour that is affected by what one believes the average behaviour of others is, i.e. the norm. The formal framework for studying interactions with such belief dependent utilities was first introduced by Geanakoplos, Pearce, and Stacchetti (1989) and then later developed by others including Battigalli and Dufwenberg (2009). As this chapter studies a purely one–shot interaction, the dynamic components of the formal theory are less applicable, however, the underlying idea of players having belief–dependent motivations is still present.

One of the implications of such motivations is the potential for a guilt averse person who experiences the negative emotion of guilt if they believe that they do not meet the expectations of others[1] (Charness & Dufwenberg, 2006). As this guilt (or expected/anticipated guilt) enters the decision makers utility function, they may behave differently to how they would if they were not guilt averse. Being more cooperative a social dilemma, for example.

Such guilt can arise regardless of whether or not behaviour is observed by others. For example, I may feel guilty if I donate less than a "suggested donation", even when no one will directly observe my donation. The emotion of shame can be distinguished from guilt by only playing a role when behaviour is observed by others (Tadelis, 2011), i.e. there is ex–post disclosure. This definition of shame gives rise to the potential for a shame averse person who experiences the negative emotion of shame if they if they do not meet the expectations of others (as in guilt aversion), but only when their behaviour is observed by others.

By varying whether and how behaviour is observed in the social dilemma of a two–player public goods game, this paper explores guilt aversion and if some people (also) exhibit shame aversion. The social dilemma studied here is an interaction amongst players that is quite different from the ones studied previously as personal and social interests are more directly at odds with each other. Namely, Tadelis (2011) uses a "noisy" trust game and Greenberg, Smeets, and Zhurakhovska (2015) use a

[1] This expectation is a second order belief with it being the belief of what others believe others will do.

sender–receiver game, both being games where any social dilemma aspect is not as clear as it is in the public goods game used here. The role of guilt and/or shame may therefore differ from what has been found in this relatively limited existing literature. With social dilemmas being present in many real world scenarios a more thorough investigation of the determinants of behaviour may reveal new and economically valuable policy recommendations. The experimental design also builds on previous work that looks at the directly elicited beliefs of subjects and how these correlate with behaviour.

A two–player, one–shot, public goods game is used with three treatments that vary the level of ex–post disclosure partners receive about decisions and outcomes (i.e. whether one's contribution decision in the public goods game will be observed by one's partner or not), along with, the equivalent ex–post disclosure received about the decisions made by partners. The experiment consists of 8 periods with stranger re–matching between each period with payment for one random, undisclosed period at the end of the experiment. This prevents reputational considerations and the ability for subjects to infer the contribution of specific partners from their payment when such disclosure is intended to be restricted.

First order beliefs (what a subject believes their partner will contribute) and second order beliefs (what a subject believes their partner expects them to contribute) are elicited directly with subjects being rewarded for correct predictions. Beliefs are elicited after each period and this allows for an investigation into if and how beliefs are shifting endogenously during the experiment. This may occur through the nature of the repeated interactions (although each time with a new subject) and/or feedback after a period if it is received. This also permits a within–subjects analysis of the role of beliefs that mitigates any potentially confounding (false) consensus effects that may be present in a between–subject analysis of beliefs (Ellingsen, Johannesson, Tjøtta, & Torsvik, 2010). It also avoids any concerns related to subjects' beliefs being elicited without informing subjects that these beliefs will subsequently be sent to another subject, a methodology which has previously been used to "induce" accurate second order beliefs (Ellingsen et al., 2010; Khalmetski, Ockenfels, & Werner, 2015; Dhami, Wei, & al Nowaihi, 2017). Any change in beliefs at an individual level could also be seen as more "natural" and externally valid when compared to other, more exogenous, mechanisms to shift beliefs as any shift is brought about through mechanisms (repeated interactions and/or feedback) that mirror the real world.

The experiment occurs in a setting where subjects meet only once and anonymity is maintained, hence, the interactions can be seen as truly one–shot and reputation or image concerns cannot play a role. These are mechanisms that can be used to help explain prosocial/cooperative behaviour but appear to have limited application to the many real world situations where there is complete anonymity or the scope of long–term reputation or image effects are limited. For example, retaliation or other consequences are not usually expected from simply saying that you are too busy to a direction asking stranger. There may however be the potential for feelings of guilt and shame if one believes that the direction asker expected one to help. In this example any feelings of shame could also depend on the number of passers-by who may also observe your interaction and if you chose to behave cooperatively or not.

This paper will first review the related literature and present a framework that builds on existing psychological game theory. This is followed by a description of the laboratory experiment designed to test for guilt and shame aversion that has been outlined above. Following this, the results are presented which suggest that at least some people are shame averse in addition to being guilt averse, the evidence for this being through a stronger response to second order beliefs when behaviour is observed ex–post. It is also found that receiving feedback on the behaviour of others works to reinforce any shift in average contribution behaviour with evidence that limiting such feedback/information can help prevent a decline in contributions. The paper concludes with a discussion of these results.

2 Literature review and theoretical framework

The monetary payoff of a public goods game with two players can be modelled as follows. Each of the players receives an endowment of $y > 0$ monetary units. They then simultaneously choose to contribute $g_i (0 \leq g_i \leq y)$ to the public good. The sum of the contributions from both players are then multiplied by a factor, $2r$, before being shared equally among the two players. The net return to an individual from contributing one unit, i.e. the marginal per capita return (MPCR), is therefore the factor, r. The monetary payoff function for player i is given by equation 1.

$$\pi_i = y - g_i + r_i(g_i + g_{j \neq i}) \tag{1}$$

By imposing the typical public goods assumption that $r < 1$, it follows that in terms of monetary payoffs, the benchmark Nash equilibrium result of zero contributions (i.e. complete free riding) and no public goods provision is obtained.

Dhami et al. (2017) build on existing psychological game theory literature to formally develop a model for a two–player public goods game that introduces psychological tendencies such as guilt aversion and surprise seeking. In their framework, a positive correlation between contributions and second order beliefs is consistent with subjects being relatively guilt averse and they find evidence for guilt aversion in aggregate.

There is also some work that has begun to look at the role of shame separately from guilt. Tadelis (2011) uses a "noisy" trust game and finds that cooperation increases significantly when there is ex–post disclosure and others will find out if a participant was truthful or not. This is the case even if the precise identity of the participants remains anonymous. In a sender–receiver game (Greenberg et al., 2015), certain ex–post disclosure is found to increase the rate of truth telling for at least some individuals.

The two–player public goods framework used by Dhami et al. (2017) embodies the social dilemma in a simple and tractable way that readily extends to more players and should therefore also offer more general insights into behaviour in social dilemmas. Their model and experiments involve ex–post disclosure, however, as argued above, guilt and guilt aversion can be seen to play a role when there is no ex–post disclosure of behaviour to others and can arise purely "internally". In line with the existing literature discussed above, any additional guilt aversion involved with ex–post disclosure can therefore be termed as shame aversion and such terminology will be used throughout this paper.

3 Experimental design and procedures

Note that the experimental instructions and the comprehension questions that participants were asked are given in Appendix B.

A variation of a standard two–player public goods game is used and at the start of each period, subjects receive an endowment of 20 tokens with the sum of contributions to the public good (described to subjects as a "project") being multiplied by 1.6 before being shared equally among the two subjects. Every token kept earns one Experimental Currency Unit (ECU) and each token contributed to the public good increases the payoff of all group members by 0.8 ECU each. The monetary payoff function for player i is therefore given by equation 2 where $g_i (0 \leq g_i \leq 20)$ is the number of tokens contributed to the public good by player i.

$$\pi_i = 20 - g_i + 0.8(g_i + g_{j \neq i}) \tag{2}$$

The experiment consists of eight periods with stranger rematching so that subjects will only ever be matched with a partner[2] for one period at most. This is reiterated to subjects in the instructions, the pre–experiment control questions, as well as, before each period. This ensures that each period can be seen as a one–shot interaction and any possible reputation effects are excluded as the aim of this paper is to look at behaviour in social dilemma situations where people remain anonymous and reputation or image concerns are absent. Subjects are aware that they will only be paid for one period, chosen at random, and they will not be told which precise period this payment relates to, thereby preventing subjects from being able to infer the exact choices of other subjects when disclosure is intended to be limited.

3.1 Treatments/roles varying ex–post disclosure

A 3 × 1 between–subject design is implemented where the level of ex–post disclosure after each period is varied across treatments. In the **"No Disclosure" (ND)** treatment (later also referred to as the ND role), participants do not receive any information on the contribution choice of their partner or the resulting payoff.

[2]Note that in the interests of neutrality; a particular players' partner is only ever referred to as the "other participant" in the running of the experiment.

The **“Asymmetric Disclosure” (AD)** treatment discloses the contribution of one subject in the pair to the other subject. The subject who receives no feedback, as in the ND treatment, will subsequently be referred to as the **Uninformed (AD–U)** subject with them taking the AD–U role. The other subject in each pairing whose contribution is not revealed, but who does receive information on what their partner contributes, is the **Informed (AD–I)** subject in the AD–I role.

The **“Full Disclosure” (FD)** treatment/role extends the disclosure received after each period to all subjects.

Treatment	Number of subjects
ND, No Disclosure	48
AD, Asymmetric Disclosure	48
FD, Full Disclosure	48
Total	144

Table 1: Summary of 3 $\times$ 1 between–subject design.

It seems intuitive that being able to observe what one’s partner contributed might give a greater potential for feeling guilt through knowing that one will later be able to make a comparison with what one’s partner contributed. Indeed, Miettinen and Suetens (2008) found that unilateral defectors feel more guilt in a prisoner’s dilemma and such an effect seems likely to be present in the very similar two–player public goods game used here. This is the reasoning behind the AD treatment as for the uninformed (AD–U) subjects, the only difference to the ND treatment is that they know that their partner will observe their contribution decision and related payoff. The AD–U subjects do not receive any kind of ex–post disclosure or feedback so there is no potential for additional guilt via the mechanism described above.

Informed subjects (AD–I) might have greater potential for feeling guilty as they will observe what their partner contributes in a similar way to those in the FD treatment. Their contribution decision is however not revealed to their partner so ex–post disclosure is still absent as it is in the ND treatment. Therefore, the only difference between these AD–I subjects and those in the FD treatment is that in the FD treatment, a subject’s partner will observe their contribution decision and related payoff.

Subjects are either assigned to being of type A or B which correspond to uninformed and informed respectively in the AD treatment. These same types are however used in all treatments in order to keep fixed any possible “in–group” or “out–group” effects related to these types (Tajfel, Billig, Bundy, & Flament, 1971; Chen & Li, 2009; Chakravarty & Fonseca, 2016). These types are announced to subjects individually on the screen and remain fixed throughout the session with subjects only ever being matched with someone of the other type and this is all common knowledge.

As discussed above, at the end of the experiment, one period is randomly selected for payment with only the payoff being declared and no additional information about which period this relates to or their partner’s decision. This is intended to maximise the relative importance of the decision in each period whilst minimising the possibility of subjects being able to infer their partner’s choice from their payoff when it is not intended.

It seems intuitive that subjects receiving ex–post disclosure might use this additional information to update their beliefs of what they expect others to do as part of some kind of learning process over the course of the experiment. Such disclosure is only present for AD–I and FD subjects and, by design, not a potential confounding factor for ND and AD–U subjects. Comparisons between the different roles do however give an insight into the role learning may play and how this compares and interacts with the other mechanisms being studied.

3.2 Beliefs

A within–subjects analysis of any changes in the first and second order beliefs over the course of the experiment and how these correlate with behaviour should offer the opportunity to explore and test the guilt/shame aversion hypotheses.

Some previous work such as that by Charness and Dufwenberg (2006); Dufwenberg, Gächter, and Hennig-Schmidt (2011) test the guilt aversion hypothesis by examining the between–subject correlation between second order beliefs and behaviour. They themselves, as well as others (Vanberg, 2008; Ellingsen et al., 2010), note that a possible confound to such an analysis is possible (false) consensus effects where people believe others think and act like themselves. Any such effect would imply that those holding higher second order beliefs would be exactly those that do in fact behave more cooperatively, i.e. in a guilt averse way. This therefore generating a positive correlation suggestive of guilt aversion that may, however, be driven primarily or solely by this consensus effect.

To address this concern, designs have been used that involve informing subjects of the elicited belief of other subjects, with these beliefs necessarily being elicited without the subjects knowing that they would be to other participants (Ellingsen et al., 2010; Khalmetski et al., 2015; Dhami et al., 2017). This "induces" an accurate second order belief, on which the effect on behaviour can be more robustly examined. Such designs can however be seen as being somewhat misleading due to the elicitation of a true belief relying on the assumption that the subject does not expect their belief to be transmitted to another subject. If this is not the case, the belief effectively becomes a form of communication which would almost certainly alter the subject's incentives. The authors themselves note that such designs "might lead to a general suspicion among participants that seemingly simple decisions may have unforeseen consequences" (Khalmetski et al., 2015), possibly distorting decisions in the rest of the experiment or subsequent experiments and hence, "contaminating" the subject pool and acting as a negative externality to some extent (Ellingsen et al., 2010; Dhami et al., 2017).

By directly eliciting first and second order beliefs with incentives for correct guesses in all periods and in all treatments [3], the experimental methodology used here does not involve such a design and is instead more similar to work by Khalmetski (2016) that tests guilt aversion with an exogenous shift in beliefs caused by changing the parameters of the experiment. The design employed here builds on this by looking at a public goods game and also offers a novel method for potentially inducing changing beliefs in order to look at the their role on behaviour in such situations. Furthermore, any change in beliefs can be seen as being more similar to real world settings. This being for the reason that any shift in beliefs can be seen as being more endogenous in the sense that there is no explicit attempt to manipulate them, they come about through feedback and/or learning over multiple periods.

The belief elicitation occurs on the same input screen as the contribution decision

[3] Beliefs relating to the subject's specific partner are elicited along with those relating to the larger pool of their eight potential partners in the session. In the following analysis, the mean of these two forms of beliefs are used. A two–sided Wilcoxon signed–rank test reports that there is no significant difference in terms of the two types of reported second order belief ($p = 0.752$) which is what the main findings relate to.

in order to ensure that the choice and beliefs reported are consistent with each other when the participant is making their choices. There is some evidence (Croson, 2000) that the act of belief elicitation can affect actions and although this design does not entirely prevent this, it does ensure consistency when playing for multiple periods as a possible situation where participants make their first choice without initially thinking about the beliefs is avoided.

3.3 Experimental procedure

The experiment was computerised, being programmed and run using z–Tree (Fischbacher, 2007). Participants were paid privately in cash immediately after the experiment using an exchange rate of 1 ECU = £0.35. Mean earnings were £9.07 with a standard deviation of £2.01, a minimum of £5.60 and a maximum of £13.30. Subjects were recruited for 60 minutes and sessions typically lasted around 50 minutes. None took longer than 60 minutes. For each of the three treatments, three sessions were conducted, each with 16 subjects, to give 48 subjects per treatment as shown in Table 1.[4] The instructions were read aloud to the room and subjects kept a copy of the instructions throughout the session. To ensure understanding, participants were given the opportunity to ask questions and they first completed a series of on–screen control questions. These questions mention no specific examples of possible contribution levels as these might otherwise act as some kind of signal or focal point. If participants answered any question incorrectly then they had to raise their hand for an experimenter to come and assist them, again ensuring that no specific behaviour is suggested. Before payment, there was a post–experiment questionnaire asking for basic demographic information such as age, nationality and gender.

Participants were recruited through the software ORSEE (Greiner, 2015) from a subject pool of over 1500 University of Exeter undergraduate students that had volunteered to participate in experiments at the Finance and Economics Experimental Laboratory at Exeter (FEELE). Students with Economics or Psychology as their main field of study were excluded from the random sample of the subject

[4] It is worth noting that each participant in the ND and AD–U roles can be seen as completely independent from each other as they never receive any information on the behaviour of others or the outcome of any period during the course of the experiment. Additionally, some Models in the results presented in Section 4.1 report standard errors at the session level and these never affect the significance of the findings presented.

pool invited to participate. This being done as there is literature indicating that Economics students may behave differently to other students and the wider population in general, either through selection effects or through studying economics itself (Bauman & Rose, 2011; Frank, Gilovich, & Regan, 1993). Psychology students were also excluded in order to avoid including other participants (Economics students already being excluded) that would be more likely to have prior experience of social dilemma games. Something which could otherwise affect the results through prior expectations and/or knowledge of the related theory adversely affecting the homogeneity of the subject pool.

As over–recruitment was used, when there were too many subjects arriving on time to a session, a random draw from an urn was used to select those to be turned away with a show up fee of £5. Participants were seated at numbered computer terminals randomly assigned by each participant drawing a number from an urn and the further computerised randomisation of player types and matching was made clear to subjects in the instructions.

The sessions were all run at the FEELE lab at the University of Exeter, UK and conducted over three days, 27th – 29th March 2017. The ordering of the treatments ensured that the sessions for each treatment were conducted at different times each day.

4 Results

In what follows there will first be a look at behaviour on a more aggregate level. This is followed by a a more detailed within–subject analysis that makes best use of the data generated by the specific experimental design used.

Finding 1: *Contributions are on average higher for those in the No Disclosure role than in any of the other roles.*

Support: Figure 1 reports the mean contribution level for each of the eight periods, by role. It shows that contributions in the first period begin at a similar level for each of the roles and are at a level broadly in the range of similar experiments where contributions typically begin around or just below 50% of the endowment (an endowment of 20 in this case). Taking these contribution levels in the first period, the difference between any combination of pairs is not significant at the 5% level when using two–sided non–parametric Mann–Whitney–Wilcoxon (MWW) tests.[5]

When taking all periods into account, similar MWW tests reveal that the only significant differences at the 5% level are between the ND (No Disclosure) role and any of the other three roles.[6] This remains the case for any combination of periods, including individual periods, from period five onwards. These results suggest

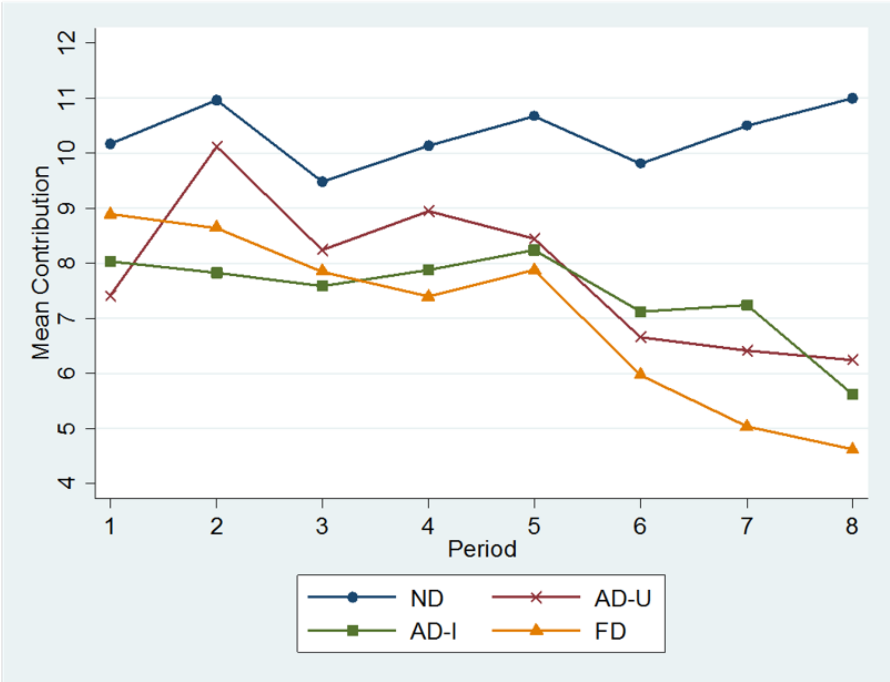

Figure 1: Mean contribution each period by role

that symmetrically limiting ex–post feedback sent to others, as well as, any such information received, can help to prevent a decline in contributions over multiple, one–shot periods with stranger rematching, as is the case in the ND treatment.

It appears that simply knowing that your current partner has been receiving ex–post feedback on the behaviour of players in the same role as yourself is enough to trigger a gradual decline in contributions. This corresponds to the AD–U (Asymmetric – Uninformed) role where players receive no additional information relative to those in the ND role. They remain just as uninformed about the behaviour of others in past periods. This indicates that a deeper analysis, including a look at

[5]ND/AD–U: $p = 0.128$; ND/AD–I: $p = 0.177$; ND/FD: $p = 0.378$; AD–U/AD–I: $p = 0.661$; AD–U/FD: $p = 0.365$; AD–I/FD: $p = 0.519$

[6]ND/AD–U: $p = 0.000$; ND/AD–I: $p = 0.000$; ND/FD: $p = 0.000$; AD–U/AD–I: $p = 0.556$; AD–U/FD: $p = 0.343$; AD–I/FD: $p = 0.932$

beliefs, is needed in order to better understand the mechanisms involved.

The similar declines in contribution levels observed in the AD–I (Asymmetric – Informed) and the FD (Full Disclosure) roles occur when players receive ex–post feedback after each period on the contribution made by their partner in that period. These players therefore do receive more information relative to the ND and AD–U players. Once again, an analysis incorporating a look at beliefs will help to identify the key mechanisms at play.

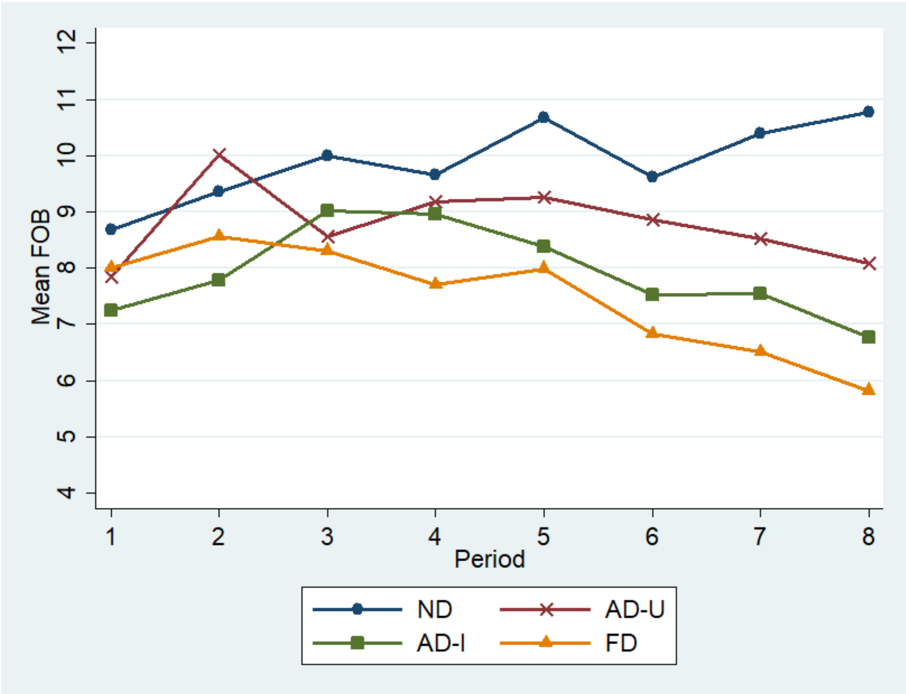

Figure 2: Mean first order belief (FOB) each period by role

Figures 2 and 3 show the evolution of first and second order beliefs over the course of the eight periods. At fist glance it appears that both sets of beliefs are very similar, and, in turn, closely resemble the actual contribution levels in each role.

This is further shown when plotting the two sets of beliefs against contributions by each role in Figures 4 and 5. Any differences between roles using such an analysis not found to be statistically significant. This is, however, not surprising as here the data has been aggregated across all participants and periods.

4.1 Within–subject analysis

In addition to looking at the data at an aggregate level, the data also allows analyses at a more individualised level. This enables a within–subject analysis regarding the

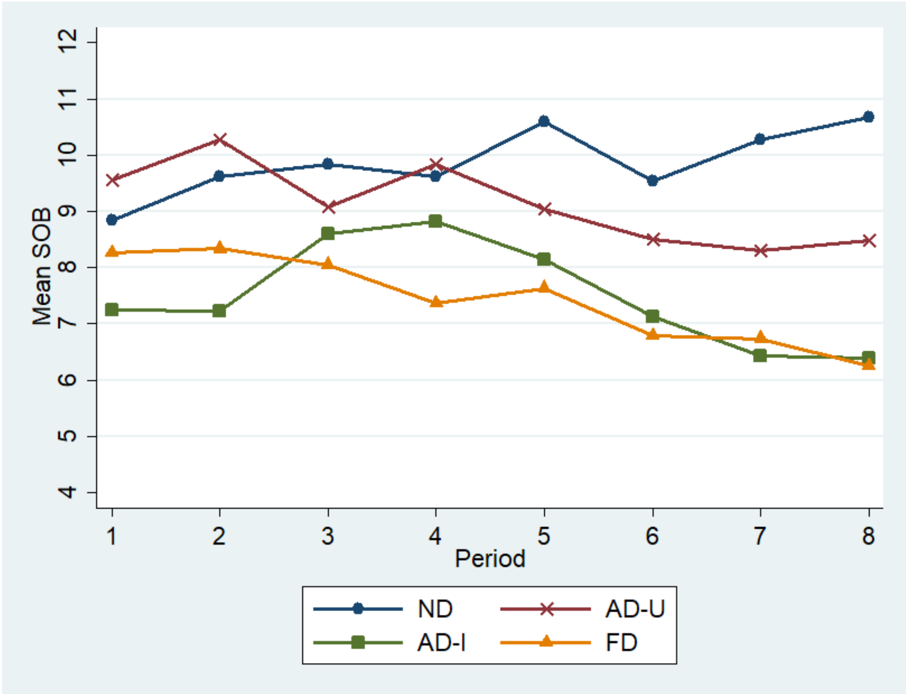

Figure 3: Mean second order belief (SOB) each period by role

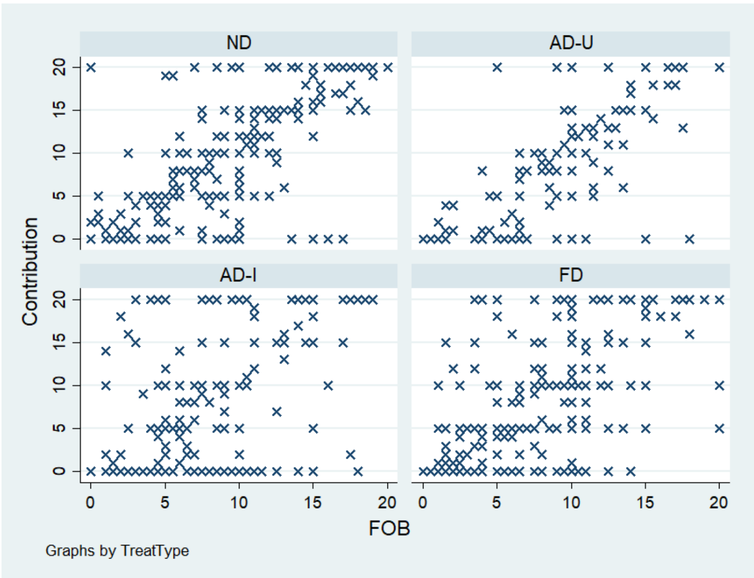

Figure 4: First order beliefs (FOB) and contributions by role

effect of beliefs on contributions, this being one of key aims of the experimental design used.

A fixed–effects ordinary least squares (OLS) estimation allows such a within–subject analysis. Controls for any subject specific characteristics, such as age or gender, are not required due to the fixed–effects analysis only taking into account variation within the set of observations from each individual participant.

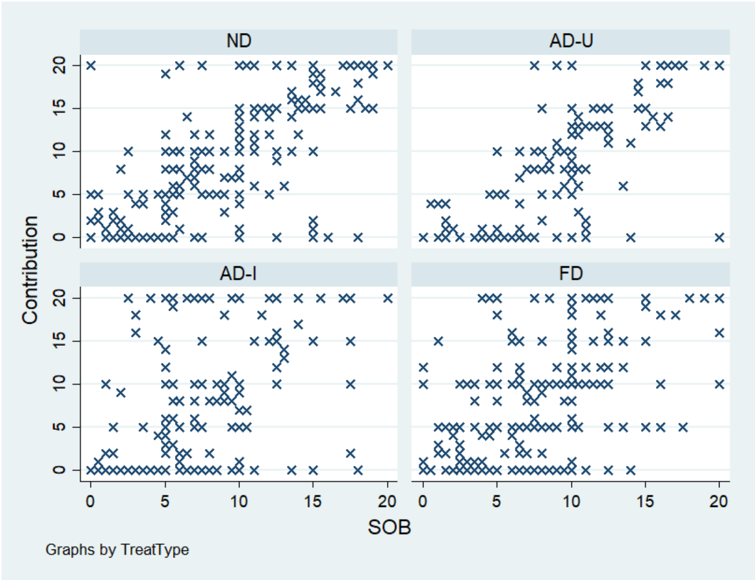

Figure 5: Second order beliefs (SOB) and contributions by role

Appendix A reports an alternative quasi–maximum likelihood (QML) estimation that attempts to mitigate any issues related to the time horizon of the data being relatively short (Hsiao, Pesaran, & Tahmiscioglu, 2002). As all the findings are still supported by this alternative estimation, the OLS estimations are presented here.

Table 2 reports the regression results when including first and second order beliefs as explanatory variables. The ND role is taken as a baseline and interaction terms are included for the types of disclosure. The two possible types of disclosure are disclosure to oneself of the outcome of each period (i.e. feedback), as in the AD–I and FD roles. The other type of disclosure is disclosure to one's partner as in the AD–U and FD roles.

Including an additional interaction term for there being both disclosure to oneself and one's partner, as there is in the FD role, does not have any significant impact with this additional term never being significant at the 5% level. As there is therefore little evidence for any additional combined interaction, such a term is omitted from the specifications presented.

Appendix A reports results interacting the relevant terms with the individual roles rather than the type of disclosure. All the findings are supported but the specification presented here is preferred as it is easier to see the effects related to varying the type of disclosure.

APPC refers to the average past partner contribution (APPC). This variable

Table 2: Beliefs and type of disclosure interactions

	(1) FE–OLS	(2) FE–OLS	(3) FE–OLS	(4) FE–OLS
FOB	0.832***	0.832***		
	(0.115)	(0.105)		
FOB × Disclosure Self	–0.565***	–0.565***		
	(0.128)	(0.108)		
FOB × Disclosure Partner	0.311***	0.311***		
	(0.118)	(0.0857)		
SOB			0.814***	0.814***
			(0.0988)	(0.0655)
SOB × Disclosure Self	-		–0.715***	–0.715***
			(0.119)	(0.0726)
SOB × Disclosure Partner			0.316***	0.316***
			(0.113)	(0.0683)
APPC × Disclosure Self	0.371***	0.371***	0.544***	0.544***
	(0.0973)	(0.0937)	(0.116)	(0.0764)
Constant	0.452	0.452	0.443	0.443
	(0.642)	(0.479)	(0.598)	(0.345)
N	1008	1008	1008	1008

Dependent variable: Contribution

Note: Robust standard errors in parentheses for Models 1 and 3

Clustered standard errors at the session level for Models 2 and 4

APPC = Average Past Partner Contribution (from the preceding periods)

All models use fixed–effects (FE) ordinary least squares (OLS)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

takes the mean of the contributions of the partners matched with in previous periods and can therefore not be calculated for the first period. This is intended to allow a more detailed look at, and incorporate into the empirical model, the effect of the additional information that some participants receive on the behaviour of others as they receive feedback over multiple periods. As would be expected, this APPC term is only significant when there is actual feedback. Hence, only one interaction term is included. Appendix A presents alternative specifications including full interactions terms.

Models 2 and 4 use the more conservative method of clustering standard errors at the session level, as opposed to robust standard errors in Models 1 and 3. It can be seen that this has relatively little impact.

Finding 2: *Guilt aversion in the absence of observability. In the ND (No Disclosure) treatment, higher (within–subject) second order beliefs are correlated with higher contribution levels.*

Support: In Table 2 it can be seen that the second order beliefs term corresponding to the ND treatment is significant at the 1% level.

The positive rather than negative coefficient is evidence for statistically significant guilt aversion at the aggregate level. Surprise/pride seeking would predict a negative coefficient (Dhami et al., 2017).

Finding 3: *Shame aversion. The positive (within–subject) response of contributions to higher second order beliefs is significantly stronger when there is ex–post disclosure to one's partner.*

Support: The positive interaction term for disclosure to one's partner is also always significant at at least the 1% level. The magnitude of this interaction term is very similar across different models, irrespective of the estimator used, whether controls for the average past partner contribution are included or whether standard errors use the more conservative clustering at the session level (see Appendix A for full results). The magnitude of this additional "shame aversion" component (when participants know that their action will be observed by their partner) is in the region of an additional 40% response to second order beliefs compared to when there is no observation. As a brief example, this would equate to around an additional 3 tokens contributed for a second order belief of 10 out of the 20 tokens.

Finding 4: *The positive (within–subject) response of contributions to higher second order beliefs is significantly reduced when there is ex–post disclosure to oneself.*

Support: The negative interaction term for disclosure to oneself is significant at the 1% level. This suggests that once there is some kind of feedback, this significantly attenuates the role of second order beliefs in explaining contributions. As a result of this, there is little evidence for second order beliefs having any impact when there is only disclosure to oneself, as in the AD–I role. As guilt aversion predicts a positive relationship, there is, therefore, limited evidence for guilt aversion in the AD–I role.

Finding 5: *Findings 2, 3 and 4 are replicated when looking at first order beliefs in place of second order beliefs.*

Support: In Table 2 it can be seen that all the relevant (interaction) terms related to first order beliefs are significant at at least the 1% level. Furthermore, the magnitudes are broadly comparable to those when second order beliefs were analysed.

This is somewhat unsurprising as first and second order beliefs are found to be very strongly correlated with each other and this finding does not discredit the earlier results. It does however show that the earlier guilt and shame aversion results do also work through first order beliefs as well as second order beliefs and differing responses to these in different experimental settings, i.e. the type of disclosure present.

It is worth noting that the significance of first order beliefs has previously been observed, with it typically being referred to and modelled as reciprocity and this being defined by earlier work on psychological game theory (Dufwenberg et al., 2011). This being different from guilt aversion which is normally seen as being related to second order beliefs.

Due the the significant correlation between first and second order beliefs discussed above, they are never included together in the Models presented here. Appendix A does however present a series of Models with them both included simultaneously along with the relevant interactions terms. Unsurprisingly, many of the terms become insignificant however this is not a concern as the high correlation between first and second order beliefs severely impacts the validity of the results of these Models. This is also an issue that others have also encountered (Dufwenberg et al., 2011).

Finding 6: *Observing the behaviour of others in previous interactions works to reinforce any shift in average behaviour.*

Support: When there is such observation by participants, the interaction term APPC $\times$ Disclosure Self is always significant at the 1% level with a positive coefficient in all Models. As this is a within–subject analysis, this has to be interpreted as the effect on contributions of a change in this average over the course of the experiment. The positive coefficient therefore implies the above finding and helps explain the observed decline in average contributions in certain roles; the presence of feedback appears to allow a kind of positive feedback mechanism to take hold. In this experiment, this appears to have worked in the direction of decreasing contributions further and faster although one could imagine this to work in the opposite direction if contributions were to follow an upward trend over the first few periods. Something that future work could potentially examine further.

There is no convincing evidence for gender or any of the other demographic data or questions included in the post experiment questionnaire interacting with the above findings. This analysis is however limited by having to use sub–sample analysis and reducing the sample sizes used.

5 Conclusion

This study has shown that in a rather general social dilemma setting (a two–person public goods game), at least some people do appear to care what others think of them, with robust evidence for "shame aversion" when people know that others who are affected monetarily by one's decision will directly observe this decision ex–post. This works through there being a stronger positive response to second order beliefs about what one believes others expect when there is such ex–post disclosure. Interestingly, the same effects are found in relation to first order beliefs although these are found to be highly correlated with each other and the key thing to note is that disclosure does have an affect and this works through beliefs, be they through first and/or second order beliefs.

This occurs in a relatively low stakes, lab based setting where there is complete anonymity and hence, no potential for any kind of reputation effects. It would be useful for future work to explore how this effect works in less restrictive settings. Exploring how this plays a role when communication is allowed, for instance.

This "shame aversion" builds on the associated hypothesis of guilt aversion which is typically related to second order beliefs and is, by definition, possible even when there is no such disclosure. This study has tested and found evidence for this guilt aversion using a within–subject design that avoids any consensus effect or experimental design concerns as it uses a more endogenous shift in beliefs over the course of the experiment. Again, the same is found when looking at first order beliefs only.

The other main finding is that personally receiving such disclosure/feedback on the behaviour of others has a significant impact with there being strong evidence that the mechanism works to reinforce any apparent trend that players observe. In the experiment conducted, this caused such greater information availability to hasten the decline in contributions. In the setting where no one received any kind of disclosure/feedback, contributions remained remarkably stable at a relatively high level. This is another aspect that it would be good for future work to explore in more detail, clarifying, for example, under what conditions limiting information about others' behaviour can help encourage prosocial behaviour.

6 Appendix A. Additional econometric analysis.

It has been shown that the often used ordinary least squares (OLS) or generalised least squares (GLS) estimators can lead to multiple issues when the time horizon is short, as is the case here (Hsiao et al., 2002). Building upon the theoretical work on dynamic fixed–effects models by Hsiao et al. (2002), quasi–maximum likelihood (QML) estimation is an alternative that can be used to avoid any biases and the following analysis uses an estimator (xtdpdqml command in Stata) developed for such short time horizon dynamic panel data (Kripfganz, 2016). As part of the estimator, the lagged dependent variable of the contribution in the previous period is always included as a regressor. As it does not seem reasonable for future beliefs to affect the initial observations, the projection option with 0 leads is used (projection, leads(0)) to use only contemporaneous values of the regressors in the initial–observations projection. The model is estimated in first differences which reduces the number of observations (by one for each subject).

The following Tables all report Models using this estimator together with the more standard OLS estimations presented in the main results sections. There is, therefore, some duplication of the data presented in Table 2 and this is done for ease of comparison between the alternative specifications.

It can be seen that after including terms for the average past partner contribution, the one period lagged contribution term becomes insignificant. This suggests that biased estimates due to endogeniety of the lagged dependent variable are less of concern in these models (Kripfganz, 2016). This being part of the justification for only presenting the OLS estimation results in Table 2.

Table 3: First order beliefs and type of disclosure interactions

	(5)	(6)	(7)	(8) FE–OLS	(9) FE–OLS
Lagged contribution	0.106*	0.0488	0.0456		
	(0.0575)	(0.0604)	(0.0613)		
FOB	0.773***	0.826***	0.814***	0.832***	0.832***
	(0.121)	(0.109)	(0.109)	(0.115)	(0.105)
FOB × Disclosure Self	–0.367***	–0.682***	–0.691***	–0.565***	–0.565***
	(0.129)	(0.139)	(0.138)	(0.128)	(0.108)
FOB × Disclosure Partner	0.312***	0.302**	0.335***	0.311***	0.311***
	(0.119)	(0.134)	(0.128)	(0.118)	(0.0857)
APPC		0.0958			
		(0.106)			
APPC × Disclosure Self		0.529***	0.689***	0.371***	0.371***
		(0.182)	(0.144)	(0.0973)	(0.0937)
APPC × Disclosure Partner		0.0957			
		(0.165)			
Constant	0.835	–1.493	–0.885	0.452	0.452
	(0.712)	(0.982)	(0.859)	(0.642)	(0.479)
N	1008	864	864	1008	1008

Dependent variable: Contribution

Note: Robust standard errors in parentheses for Models 5 to 8

Clustered standard errors at the session level for Model 10

APPC = Average Past Partner Contribution (from the preceding periods)

Models 5 to 7 use fixed–effects (FE) quasi–maximum likelihood linear dynamic panel data estimation, see body of the text for more details

Models 8 and 9 use FE ordinary least squares (OLS)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Second order beliefs and type of disclosure interactions

	(10)	(11)	(12)	(13) FE–OLS	(14) FE–OLS
Lagged contribution	0.157***	0.0682	0.0670		
	(0.0601)	(0.0590)	(0.0593)		
SOB	0.723***	0.831***	0.831***	0.814***	0.814***
	(0.0996)	(0.103)	(0.102)	(0.0988)	(0.0655)
SOB × Disclosure Self	–0.407***	–0.893***	–0.902***	–0.715***	–0.715***
	(0.112)	(0.136)	(0.137)	(0.119)	(0.0726)
SOB × Disclosure Partner	0.315***	0.348***	0.361***	0.316***	0.316***
	(0.110)	(0.127)	(0.125)	(0.113)	(0.0683)
APPC		0.0255			
		(0.105)			
APPC × Disclosure Self		0.864***	0.916***	0.544***	0.544***
		(0.172)	(0.161)	(0.116)	(0.0764)
APPC × Disclosure Partner		0.0380			
		(0.148)			
Constant	0.996	–1.619*	–1.458*	0.443	0.443
	(0.639)	(0.981)	(0.876)	(0.598)	(0.345)
N	1008	864	864	1008	1008

Dependent variable: Contribution

Note: Robust standard errors in parentheses for Models 10 to 13

Clustered standard errors at the session level for Model 14

APPC = Average Past Partner Contribution (from the preceding periods)

Models 10 to 12 use fixed–effects (FE) quasi–maximum likelihood linear dynamic panel data estimation, see body of the text for more details

Models 13 and 14 use FE ordinary least squares (OLS)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Second order beliefs and role interactions

	(15)	(16)	(17)	(18) FE–OLS	(19) FE–OLS
Lagged contribution	0.160***	0.0661	0.0662		
	(0.0601)	(0.0597)	(0.0596)		
SOB	0.704***	0.801***	0.799***	0.799***	0.799***
	(0.112)	(0.115)	(0.114)	(0.114)	(0.0731)
SOB × AD–U	0.438**	0.423*	0.450*	0.366**	0.366***
	(0.181)	(0.236)	(0.240)	(0.171)	(0.0832)
SOB × AD–I	–0.338**	–0.890***	–0.889***	–0.773***	–0.773***
	(0.149)	(0.170)	(0.169)	(0.168)	(0.190)
SOB × FD	–0.107	–0.507***	–0.506***	–0.365**	–0.365***
	(0.134)	(0.153)	(0.152)	(0.146)	(0.0935)
APPC		–0.0247			
		(0.113)			
APPC × AD–U		0.189			
		(0.142)			
APPC × AD–I		1.043***	1.018***	0.702***	0.702**
		(0.252)	(0.224)	(0.196)	(0.233)
APPC × FD		0.892***	0.867***	0.478***	0.478***
		(0.232)	(0.206)	(0.138)	(0.0222)
Constant	0.907	–1.524	–1.429	0.442	0.442
	(0.648)	(0.998)	(0.876)	(0.601)	(0.328)
N	1008	864	864	1008	1008

Dependent variable: Contribution

Note: Robust standard errors in parentheses for Models 15 to 18

Clustered standard errors at the session level for Model 19

APPC = Average Past Partner Contribution (from the preceding periods)

Models 15 to 17 use fixed–effects (FE) quasi–maximum likelihood linear dynamic panel data estimation, see body of the text for more details

Models 18 and 19 use FE ordinary least squares (OLS)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: First order beliefs and role interactions

	(20)	(21)	(22)	(23) FE–OLS	(24) FE–OLS
Lagged contribution	0.108^{*}	0.0461	0.0450		
	(0.0573)	(0.0610)	(0.0617)		
FOB	0.782^{***}	0.831^{***}	0.822^{***}	0.829^{***}	0.829^{***}
	(0.138)	(0.118)	(0.118)	(0.132)	(0.126)
FOB × AD–U	0.340	0.240	0.271	0.321	0.321^{**}
	(0.232)	(0.268)	(0.284)	(0.211)	(0.134)
FOB × AD–I	-0.361^{**}	-0.795^{***}	-0.787^{***}	-0.663^{***}	-0.663^{***}
	(0.169)	(0.162)	(0.162)	(0.170)	(0.162)
FOB × FD	–0.0750	-0.323^{**}	-0.314^{*}	–0.205	–0.205
	(0.164)	(0.162)	(0.163)	(0.165)	(0.169)
APPC		0.0262			
		(0.112)			
APPC × AD–U		0.303^{*}			
		(0.184)			
APPC × AD–I		0.855^{***}	0.881^{***}	0.571^{***}	0.571^{**}
		(0.259)	(0.233)	(0.185)	(0.184)
APPC × FD		0.560^{***}	0.587^{***}	0.271^{**}	0.271^{***}
		(0.204)	(0.174)	(0.106)	(0.0771)
Constant	0.738	–1.352	–0.830	0.464	0.464
	(0.712)	(0.990)	(0.861)	(0.639)	(0.449)
N	1008	864	864	1008	1008

Dependent variable: Contribution

Note: Robust standard errors in parentheses for Models 20 to 23

Clustered standard errors at the session level for Model 24

APPC = Average Past Partner Contribution (from the preceding periods)

Models 20 to 22 use fixed–effects (FE) quasi–maximum likelihood linear dynamic panel data estimation, see body of the text for more details

Models 23 and 24 use FE ordinary least squares (OLS)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: First and second order beliefs and type of disclosure interactions

	(25)	(26)	(27)	(28) FE–OLS	(29) FE–OLS
Lagged contribution	0.110**	0.0456	0.0424		
	(0.0558)	(0.0581)	(0.0584)		
FOB	0.614**	0.425**	0.405**	0.514**	0.514**
	(0.258)	(0.174)	(0.182)	(0.239)	(0.198)
FOB × Disclosure Self	-0.0106	-0.0152	0.00254	-0.0227	-0.0227
	(0.272)	(0.206)	(0.211)	(0.263)	(0.184)
FOB × Disclosure Partner	0.0564	0.0776	0.0834	0.0966	0.0966
	(0.239)	(0.181)	(0.179)	(0.227)	(0.177)
SOB	0.194	0.447***	0.464***	0.368*	0.368**
	(0.226)	(0.161)	(0.168)	(0.214)	(0.147)
SOB × Disclosure Self	-0.424*	-0.833***	-0.852***	-0.646***	-0.646***
	(0.243)	(0.189)	(0.200)	(0.242)	(0.118)
SOB × Disclosure Partner	0.298	0.331**	0.334**	0.242	0.242*
	(0.222)	(0.160)	(0.163)	(0.201)	(0.125)
APPC		0.0606		-0.0179	-0.0179
		(0.101)		(0.0680)	(0.0682)
APPC × Disclosure Self		0.660***	0.729***	0.402***	0.402***
		(0.167)	(0.149)	(0.126)	(0.112)
APPC × Disclosure Partner		0.0116		0.0270	0.0270
		(0.145)		(0.120)	(0.0882)
Constant	0.569	-1.578*	-1.252	0.195	0.195
	(0.681)	(0.943)	(0.831)	(0.656)	(0.615)
N	1008	864	864	1008	1008

Dependent variable: Contribution

Note: Robust standard errors in parentheses for Models 25 to 28

Clustered standard errors at the session level for Model 29

APPC = Average Past Partner Contribution (from the preceding periods)

Models 25 to 27 use fixed–effects (FE) quasi–maximum likelihood linear dynamic panel data estimation, see body of the text for more details

Models 28 and 29 use FE ordinary least squares (OLS)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7 Appendix B. Experiment Instructions

Welcome to the FEELE Laboratory. This is an experiment in decision making with funding provided by the University of Exeter. You can earn money depending on the decisions made by you and other participants today.

Please remain silent during the experiment and switch off mobile phones. If you have questions or require assistance at any point during the experiment, please raise your hand and an experimenter will come to you. Do not ask any questions out loud. Please read these instructions carefully as they are read aloud to the room. There will then be an opportunity to ask any clarifying questions and each of you will answer some on–screen questions to check understanding of the instructions.

Payment and structure

In the instructions your earnings are referred to in terms of Experimental Currency Units (ECUs). At the end of the session the total amount of ECUs that you have earned will be converted to Pounds at the rate, 1 ECU = £0.35. You will then be paid privately in cash. It is important to note that everyone will remain anonymous throughout and after the experiment.

You are part of a group of 16 participants. The experiment consists of 8 rounds. Before the experiment begins, everyone in the room will be randomly assigned to be either a participant of **type A** or **type B** for the entire experiment. There will be 8 people assigned to each type. You will be informed of your type on the screen. In each round, each type A participant interacts with a type B participant and each type B participant interacts with a type A participant. At the end of the experiment, for each participant the computer will randomly select one round that will be used to calculate earnings. You will be informed on–screen of your earnings but not which exact round has been chosen.

In each round you will be matched with one other participant who you have not been matched with before and will not be matched with again in subsequent rounds.

Decisions in each round

At the start of each round, both you and the other participant who you are matched with have **20 tokens each.** This is called your endowment. Each of you must decide how to use your endowment. You have to decide how many of the 20 tokens you want to contribute to a project and how many you want to keep for yourself. The other participant makes the same decision.

Every token that you keep for yourself earns you 1 ECU.

For the tokens contributed to the project, the following happens. The contributions from you and the other participant you are matched with are added together to give the total contribution. This total will then be **multiplied by 1.6 and this amount will be divided equally among the two of you.** Each of you therefore receives 0.8 ECUs for each token either of you contributes to the project and both of you receive the same income from the project.

Your total income in ECUs is therefore:

(20 – tokens contributed to the project by you) + (1.6/2) x (tokens contributed to the project by you + tokens contributed to the project by the other participant)

You will be presented with an input screen similar to the above where you type in the number of tokens that you would like to contribute to the project. Your decision can be changed at any point until you click the OK button. You cannot change your decision after clicking the OK button.

On the same input screen you are also asked for four predictions. The first prediction is what you think **the other participant that you are matched with for that round will contribute to the project.**

The second prediction is what you think **the other participant thinks you will contribute.** In other words, you are asked to predict the other participant's first prediction.

The third prediction is what you think **the average (mean) contribution of the 8 other participants that you could possibly be matched with** will be that round (to the nearest whole number). These participants being all of those that

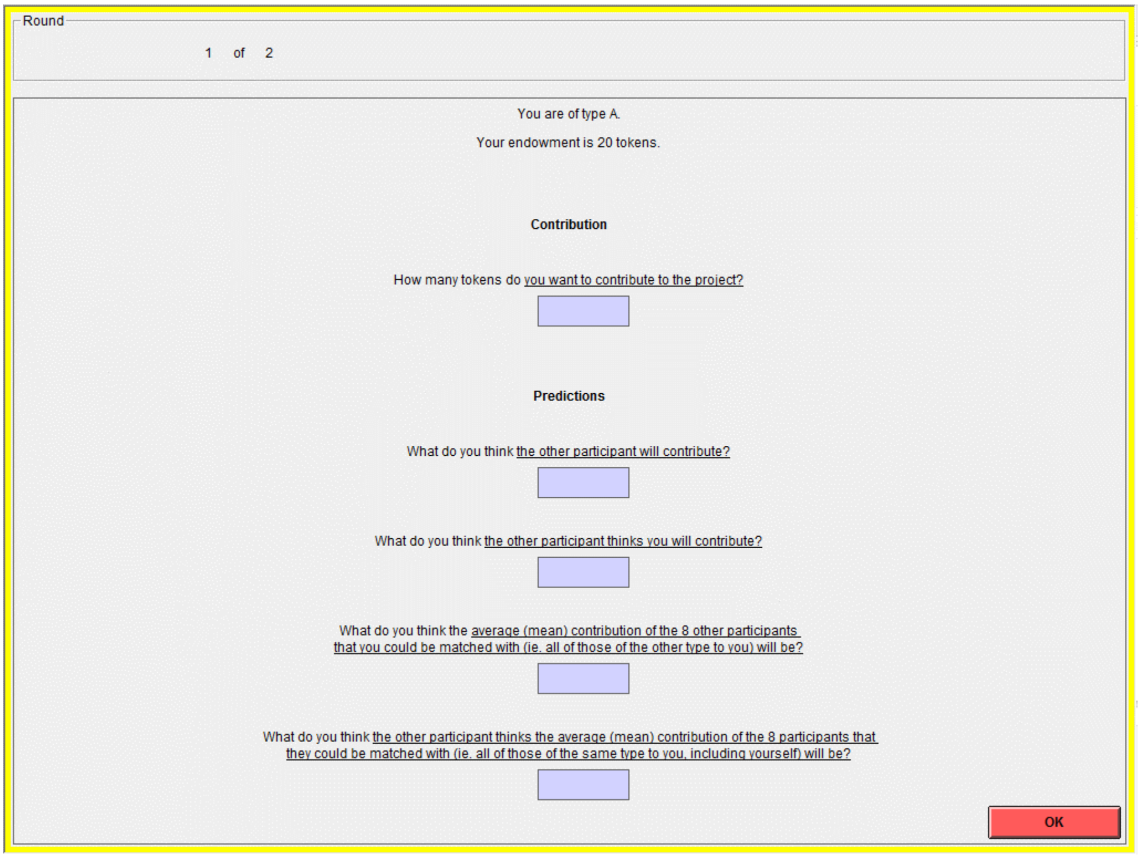

Figure 6: Screenshot of the input screen.

are of the other type to you and include the other participant that you are actually matched with.

The fourth prediction is what you think **the other participant thinks the average (mean) contribution of the 8 participants that they could be matched with** (of which you are one) will be. Here the relevant participants are all of those of the same type to you, including yourself. In other words, you are asked to predict the other participant's third prediction. If that round is selected for payment, you will receive a bonus of 2 ECUs for each correct prediction.

The following paragraph differs for the three treatments; ND (No Disclosure), AD (Asymmetric Disclosure) and FD (Full Disclosure)

ND – After the contributions and predictions have been made by both of you, that round ends. **Neither you nor the other participant will find out what choices or predictions the other made.**

AD – After the contributions and predictions have been made by both of you, type B participants will be informed of the contributions and their resulting income.

Type A participants will receive no such information. In other words, the **type A participant's contribution will be revealed to the other participant but the type B participant's contribution will never be revealed to the other participant.**

FD – After the contributions and predictions have been made by both of you, **both of you will be informed of each other's contributions and your resulting income.**

Everyone will then begin a new round after being matched with a new participant. As mentioned before, the participant you are matched with will be someone that you have never been matched with previously and will not be matched with again in subsequent rounds.

After the 8 rounds and the earnings announcement, there will be a short questionnaire whilst payments are being organised. You will then be called to come forward one by one to collect your cash payment in private.

Are there any questions at this point? There will be some on–screen questions to check your understanding of the rules of today's experiment. If you get any wrong you will be informed on the screen. Please raise your hand so an experimenter can come to you and explain the right answer. Also please raise your hand if you have any other questions or anything is unclear.

7.1 Control questions and answers (correct in bold)

1. How many rounds will the experiment run for? **8**.

2. Is there any chance of being matched with the same person for more than one round? Yes, **No**.

3. Will you find out which exact round is used for payment? Yes, **No**.

4. Does your contribution to the project affect the other participant's earnings? **Yes**, No.

5. Is the income from the project shared equally between the two of you? **Yes**, No.

6. By what factor are the total contributions to the project multiplied before being shared equally between the two of you? **1.6**.

7. Will the other participant ever find out what you contributed? ND – **No**, FD – **Yes**, No; AD – Yes, No, **Only if I am of type A**, Only if I am of type B.
8. When will you get a bonus? If your first prediction is equal to: **The other participant's contribution**, The average contribution of participants that I could be matched with, The other participant's first prediction, The other participant's third prediction.
9. When will you get a bonus? If your second prediction is equal to: The other participant's contribution, The average contribution of participants that I could be matched with, **The other participant's first prediction**, The other participant's third prediction.
10. When will you get a bonus? If your third prediction is equal to: The other participant's contribution, **The average contribution of participants that I could be matched with**, The other participant's first prediction, The other participant's third prediction.
11. When will you get a bonus? If your fourth prediction is equal to: The other participant's contribution, The average contribution of participants that I could be matched with, The other participant's first prediction, **The other participant's third prediction**.

References

Battigalli, P., & Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, *144*(1), 1–35.

Bauman, Y., & Rose, E. (2011). Selection or indoctrination: Why do economics students donate less than the rest? *Journal of Economic Behavior & Organization*, *79*(3), 318–327.

Chakravarty, S., & Fonseca, M. (2016). Discrimination via exclusion: An experiment on group identity and club goods. *Journal of Public Economic Theory*.

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*(74), 1579–1601.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *The American Economic Review*, *99*(1), 431–457.

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

Dhami, S., Wei, M., & al Nowaihi, A. (2017). Public goods games and psychological utility: Theory and evidence. *Journal of Economic Behavior & Organization*.

Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, *73*(2), 459–478.

Ellingsen, T., Johannesson, M., Tjøtta, S., & Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, *68*(1), 95–107.

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. doi: 10.1007/s10683-006-9159-4

Frank, R., Gilovich, T., & Regan, D. (1993). Does studying economics inhibit cooperation? *The Journal of Economic Perspectives*, *7*(2), 159–171.

Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, *1*(1), 60–79.

Greenberg, A., Smeets, P., & Zhurakhovska, L. (2015). Promoting Truthful Communication Through Ex-Post Disclosure. *Available at SSRN*.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 114–125.

Hsiao, C., Pesaran, M., & Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics*, *109*(1), 107–150.

Khalmetski, K. (2016). Testing guilt aversion with an exogenous shift in beliefs. *Games and Economic Behavior*, *97*, 110–119.

Khalmetski, K., Ockenfels, A., & Werner, P. (2015). Surprising gifts: Theory and laboratory evidence. *Journal of Economic Theory*, *159*, 163–208.

Kripfganz, S. (2016). Quasi–maximum likelihood estimation of linear dynamic short-T panel-data models. *Stata Journal*, *16*(4), 1013–1028.

Miettinen, T., & Suetens, S. (2008). Communication and Guilt in a Prisoner's Dilemma. *Journal of Conflict Resolution*, *52*(6), 945–960.

Tadelis, S. (2011). The power of shame and the rationality of trust. *Haas School of Business Working Paper*.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–

178.
Vanberg, C. (2008). Why do people keep their promises? an experimental test of two explanations. *Econometrica*, *76*(6), 1467–1480.