Testing for exposure effects and guilt aversion in a public goods game

Hannes Titeca*
University of Exeter

December 12, 2020

Abstract

Exposure effects are related to knowing that one's choice will be observed by others or that you will observe the choice of others ex–post. For example, a "shame averse" individual would contribute more to a public good when they know that they will be observed. This is distinct from guilt aversion where an individual experiences disutility simply from failing to meet the expectations of others. The guilt aversion hypothesis therefore predicts a positive correlation between cooperative behaviour and second order beliefs (what one believes others expect one to do). A one–shot, two–player public goods game is used to test predictions using a 3×2 between–subjects design. The experiment is conducted online using a sample more representative of the general population than typical lab studies. To avoid any potentially confounding consensus effects in the analysis, beliefs are exogenously manipulated at the treatment level through either a low or high contribution norm signal presented to subjects at the point they make their choice. It is found that there are significantly fewer contributions at zero (free–riding) when it is known that one will be observed ex–post, i.e. "shame aversion". The level of contributions at an aggregate level is however relatively unaffected. Additionally, varying the norm signal does indeed shift beliefs and contributions in the direction of the signal, even when one's choice will not be observed. This providing strong and robust evidence in support of the guilt aversion hypothesis.

Keywords: exposure effects, ex-post disclosure, social norms, guilt aversion, shame aversion, public goods game, online experiment, 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

A consistent finding in experimental economics is that of behaviour which goes against the individual incentive to maximise monetary payoff, even in a one–shot setting where individuals interact with each other only once. In social dilemmas where there is a direct conflict between individual and group incentives (at least in monetary terms), this corresponds to behaviour that increases the payoff of others at the expense of one's own payoff. One explanation for such cooperative behaviour[1] is belief–based guilt aversion where a guilt averse person experiences guilt if they believe that they do not meet the expectations of others (Charness & Dufwenberg, 2006).

It is important to consider a distinction between guilt that can be felt irrespective of whether behaviour is observed ex–post and guilt that can be felt only when observed by others ex–post. There are many settings where such a distinction between different types of guilt can be clearly seen. One example being that of litter dropping where guilt may be felt when dropping litter, even when not observed. If, however, there are onlookers, then it is clear that additional negative emotions may come into play, even if there is no expectation of any longer term consequences (punishment, social exclusion etc.). This disutility, which is only possible when observed ex–post, would be considered as "guilt from blame" in the work by Battigalli and Dufwenberg (2007). However, to make the distinction with (simple) guilt clearer, it is the term, "shame", that this paper will use to describe any negative emotions arising specifically from being observed by others. "Guilt" and "guilt aversion" refer to the negative emotions from not meeting an expectation, irrespective of whether behaviour is observed by others or not. This is also consistent with related psychology literature and some existing experimental work such as Tadelis (2011).

The public goods game is often used to model and study social dilemmas and is what this paper will use to explore these ideas of guilt and shame. In a public goods game each individual chooses how many tokens to contribute from an endowment. The sum of the tokens contributed by all individuals is then multiplied by a factor greater than one but less than the number of individuals. The resulting amount of tokens is then divided equally amongst the individuals who are left with this

[1] Often alternatively referred to as prosocial behaviour

amount in addition to the tokens they did not contribute from their endowment. It is clear that in terms of an individual's private incentive to maximise their final amount of tokens, they should contribute nothing and free–ride completely. Indeed, such behaviour by all individuals is the unique Nash equilibrium of the game when taking the number of final individual tokens as the relevant payoff. If one would want to maximise the total number of tokens across all the individuals, i.e. follow the group incentive, then one would contribute their entire endowment.

The formal framework for studying interactions with 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 paper studies a purely one–shot interaction, the dynamic components of the formal theory are less applicable, however, the central notion of players potentially having belief–dependent motivations is what will be used throughout this paper.

Some studies have looked experimentally at the role of exposure using a version of the trust mini game introduced by Charness and Dufwenberg (2006). Both Tadelis (2011) and Bracht and Regner (2013) find that the proportion of recipients choosing roll (which can be seen as the cooperative choice) increases significantly when there is exposure. Andrighetto, Grieco, and Tummolini (2015) do however not replicate this result.

In a sender–receiver game, Greenberg, Smeets, and Zhurakhovska (2015) find that certain ex–post disclosure increases the rate of truth telling for at least some individuals.

This paper complements these studies by instead using a social dilemma where personal and social interests are more directly at odds with each other. More specifically, a two–player, one–shot, public goods game is used with a 3×2 between subject design. One treatment arm varies exposure and whether subjects are observed ex–post and/or will receive ex–post feedback (i.e. observe the action of the other player).

The other treatment arm refers to whether it is a low or high contribution norm signal that is presented to subjects at the time they make their contribution choice. This is used to exogenously manipulate beliefs and allow a rigorous test of hypotheses related to correlations between beliefs and behaviour.

One of these is the guilt aversion hypothesis which predicts a positive correlation between second order beliefs and behaviour. Previous work such as that by Charness and Dufwenberg (2006) and Dufwenberg, Gächter, and Hennig-Schmidt (2011) test this hypothesis by examining the between–subject correlation between elicited beliefs and behaviour. They themselves, as well as others (Vanberg, 2008; Ellingsen, Johannesson, Tjøtta, & Torsvik, 2010), note that a possible confound to such an analysis is the (false) consensus effect where people tend to believe others think and act like themselves (Ross, Greene, & House, 1977; Ellingsen et al., 2010). This effect would mean that those holding higher second order beliefs (and first order beliefs) may be exactly those that do in fact behave more cooperatively. Thus generating a positive correlation in support of guilt aversion that may 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 sent (Ellingsen et al., 2010; Khalmetski, Ockenfels, & Werner, 2015; Dhami, Wei, & al Nowaihi, 2017). This does indeed "induce" an accurate second order belief, on which the effect on behaviour can be more robustly examined. However, such designs can 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, then 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. This may also extend to subsequent experiments that the subject is part of, "contaminating" the subject pool and acting as a negative externality to some extent (Ellingsen et al., 2010; Dhami et al., 2017).

By exogenously manipulating beliefs at a treatment level, the design used in this study eliminates the potential for consensus effects whilst not relying on a design that uses induced beliefs.

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 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 / image effects are limited. For example; retaliation or other consequences are not usually expected from dropping litter (to continue with the example given previously).

2 Experimental design and procedures[2]

A variation of a standard two–player public goods game is used. In a pairing of two players, each receives an endowment of 20 tokens from which they contribute to the public good (described to subjects as a "project"). The sum of contributions to the public good is multiplied by 1.6 before being shared equally among the two players as Experimental Currency Units (ECUs). Each token contributed to the public good therefore increases the payoff to both by 0.8 ECU each. Every token kept earns one ECU. The monetary payoff function in terms of ECUs for player i is therefore given by equation 1 where $g_i(0 \leq g_i \leq 20)$ is the number of tokens contributed to the public good by player i. ECUs were converted into Pounds at a rate of 1 ECU = £0.05.

$$\pi_i = 20 - g_i + 0.8(g_i + g_{j \neq i}) \tag{1}$$

The experiment consists of a single round of the above game with each subject matched to another subject recruited from the same online subject pool (explained in more detail below). This partner always remains anonymous.[3] The instructions make clear that they will only interact with this other subject once. This ensures that the round is seen as a one–shot interaction and any possible reputation effects are excluded. Subjects are aware that they will only be paid based on the outcome of the public goods game with 50% probability and with 50% probability they will be paid based on a random draw. This draw is from all possible amounts between and including £0.80 and £1.80, in one pence (£0.01) increments. These minimum and maximum amounts correspond to the lowest and highest possible payoffs from the public goods game. There is a new draw for each subject and subjects are, in general (with exceptions for some treatments/roles as outlined in the next section), not made aware if they are paid based on the game payoff or the random draw. This, and the fact that any possible game payoff is able to be drawn in the random draw, means that subjects are not able to infer their partners' choice in the game with certainty (hence, limiting disclosure/observabiity) when it is not intended.

[2]Note that the experimental instructions are given in Appendix A. The comprehension questions that were required to have been correctly answered before making choices (multiple attempts allowed) are given in Appendix B.

[3]In the interests of neutrality, a particular players' partner is only ever referred to as the "other participant" in the running of the experiment.

		Social norm signal	
		Low	High
Exposure	No Disclosure	ND–L (43)	ND–H (49)
	Asymmetric Disclosure	AD–L (91)	AD–H (76)
	Full Disclosure	FD–L (58)	FD–H (44)

Table 1: Summary of 3 × 2 between–subject design. Number of subjects in parentheses.

2.1 Treatments

A 3 × 2 between–subject design is implemented where the level and direction of ex–post disclosure (exposure) is varied across treatments, as well as, which of two forms of descriptive norm signal is provided to the subject. Each of the six treatments is chosen at random for each paring of two subjects. As can be seen in Table 1, a higher weight of the Asymmetric Disclosure (AD) treatments being selected was used. This was done as for some of the analyses it is required to separate these treatments into two different roles, one for each of the two subjects in a paring. For these analyses, this therefore allows the number of subjects in each comparison group to be more balanced than they would otherwise be.

Some of the treatments have an odd number of subjects due to the matching procedure that was used. As the experiment was run online it was expected that some subjects would start the experiment but not complete it entirely. It was for this reason that the matching of subjects into pairs was only done sequentially after subjects had made all the required choices [4]. Data is retained for the small number of subjects who did not complete the entire experiment (and hence were not matched to another subject) but still made a contribution choice in the public goods game and these observations are still used in the analysis and included in the totals in Table 1.

2.2 Treatments varying exposure

In the **No Disclosure (ND)** treatments participants do not receive any information on the contribution choice of their partner or the resulting payoff. The **Asymmetric**

[4]This is of no consequence to participants as they are only ever presented with information about their partner's choices (and then only in certain roles/treatments) at the end of the experiment after all choices have been made.

Disclosure (AD) treatments disclose 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. 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. The **Full Disclosure (FD)** treatments extend the disclosure received to all subjects, i.e. each subject knows that they will both find out what the other did.

As in the AD treatments there are two different roles that subjects can play (with this being fixed during the experiment), AD–U and AD–I, the rest of the paper will also refer to the ND and FD treatments by the ND and FD roles that subjects play. These four different roles allow two different potential effects to be investigated; the effect from being observed by others ex–post and the effect from observing others ex–post through receiving feedback.

The first is related to any emotions associated with being observed. These could be negative, as with shame, or potentially positive, as with surprise seeking where positive utility is derived from exceeding the expectation of others (Dhami et al., 2017; Khalmetski et al., 2015).

The second is related to any emotions through knowing that it will later be possible to make a comparison with what one's partner contributed. This could result in greater potential for negative emotions if, in hindsight, one would make a different contribution choice, i.e. regret. This would be the case for a conditional cooperator who would ideally want to match their partner's contribution. This potential or anticipated emotion might therefore influence behaviour in some way.

These considerations are the reasoning behind the AD treatments as for the uninformed (AD–U) subjects, the only difference to the ND treatments is that they know that their partner will observe their contribution decision. The AD–U subjects do not receive any kind of ex–post disclosure or feedback themselves so there is no potential for any additional guilt, regret etc. via the mechanisms described above.

Informed subjects (AD–I) may have a greater potential for feeling guilty as they will observe what their partner contributes in a similar way to those in the FD treatments. Their contribution decision is however not revealed to their partner so ex–post disclosure is still absent as it is in the ND treatments. Therefore, the only

difference between these AD–I subjects and those in the FD treatments is that in the FD treatments, a subject's partner will observe their contribution decision.

2.3 Treatments varying the exogenous social norm signal

To robustly look at the relationship between beliefs and behaviour, beliefs are exogenously manipulated by providing subjects with one of two versions of a descriptive norm signal. Henceforth, this will be referred to as the "signal".

This takes the form of a line on the decision screen stating: "The average contribution in a previous similar experiment was 5." or "...was 15." With each subject having 20 tokens available to contribute the **Low signal** treatment with a contribution of 5 corresponds to a contribution rate of 25% and the **High signal** treatment with a contribution of 15 corresponds to a contribution rate of 75%. Each treatment occurs with equal probability and is determined randomly for each pairing of two subjects.

Subjects are not made aware of the two treatments and it is made clear that the signal is shown to both subjects in a pairing. The instructions specify that the signal is the "average contribution choice made by participants in a similar experiment with the same payment structure." [5]

The signal taking the form of a common descriptive norm would therefore be expected to shift first, second and possibly higher order beliefs for at least some subjects towards the signal value. Hence, higher beliefs would be expected in the high signal treatment and this is tested through the direct elicitation of first and second order beliefs with incentives for correct predictions.

As outlined before, the manipulation of beliefs in this study does not involve an "induced" belief 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.

[5] Although not explained to subjects, the level of 5 tokens is the mode and median of contributions between and including 1 and 9 in a previous experiment that had the same public goods game and three exposure treatments as this experiment. 15 tokens was the mode and median of of contributions between and including 11 and 19.

2.4 Experimental procedure

The experiment was computerised, being programmed and run using oTree (Chen, Schonger, & Wickens, 2016). The experiment was run online with the experiment hosted on Heroku (https://www.heroku.com) and participants were recruited and paid privately using Prolific (https://prolific.ac).

The exchange rate used was 1 ECU = £0.05. Average earnings were £1.34 with a standard deviation of £0.25, a minimum of £0.80 and a maximum of £1.87. The average time taken was 510 seconds (8.5 minutes) with a standard deviation of 204 seconds. This equates to an average payment per hour of £9.46, broadly in line with typical lab and online experiments.

Subjects were recruited and matched sequentially with other subjects as they were recruited in real–time. Allocation to each of the six treatments was random but fixed for each pairing of two subjects.

The instructions were shown to subjects before a series of comprehension questions that had to be correctly answered before moving on to the decision screen. Multiple attempts were allowed. A copy of the instructions was available to players on the same screen as the comprehension questions and decision. The comprehension questions mention no specific examples of possible contribution levels as these might otherwise act as some kind of signal or focal point.

Self reported demographic information such as age, gender, current country of residence, nationality and student/employment status is recorded before recruitment by the Prolific platform and hence available for analysis without having to be asked during the experiment.

At the time the first subjects were recruited there were over 38,000 people registered on the recruitment platform and no specific screening was applied so subjects were invited randomly from this pool until the desired number of participants was reached. Recruitment was carried out during regular UK working hours and the largest group was currently resident in the UK although this was still less than half of the achieved sample. The majority of the sample were not students and the majority were either working full or part time. The average age was 30.1 with a standard deviation of 10.00, a minimum of 18 (imposed by the recruitment platform) and a maximum of 66.

The recruitment and experiment was carried out on three different days; Wednesday 25th July 2018, Thursday 31st January 2019 and Monday 4th February 2019.

3 Results

3.1 Exposure effects

Finding 1 (shame aversion): *Ex–post disclosure significantly reduces contributions of zero (complete free–riding).*

Support: Figure 1 reports the proportion of complete free–riding (zero contribution) by ex–post disclosure. As a reminder, ex–post disclosure is present in the AD–U / FD roles and there is no ex–post disclosure in the ND / AD–I roles. It shows a significant difference with less than half the amount of free–riding when there is disclosure and subjects know that their partner will observe their contribution ex–post, compared to when there is no such disclosure. This difference being significant at the 1% level when using a two–sided Fisher's exact test ($p = 0.002$).

In Figure 1 and subsequent figures the grey bars show the 95% confidence intervals computed/generated using the CIBAR package for STATA (Staudt, 2019).

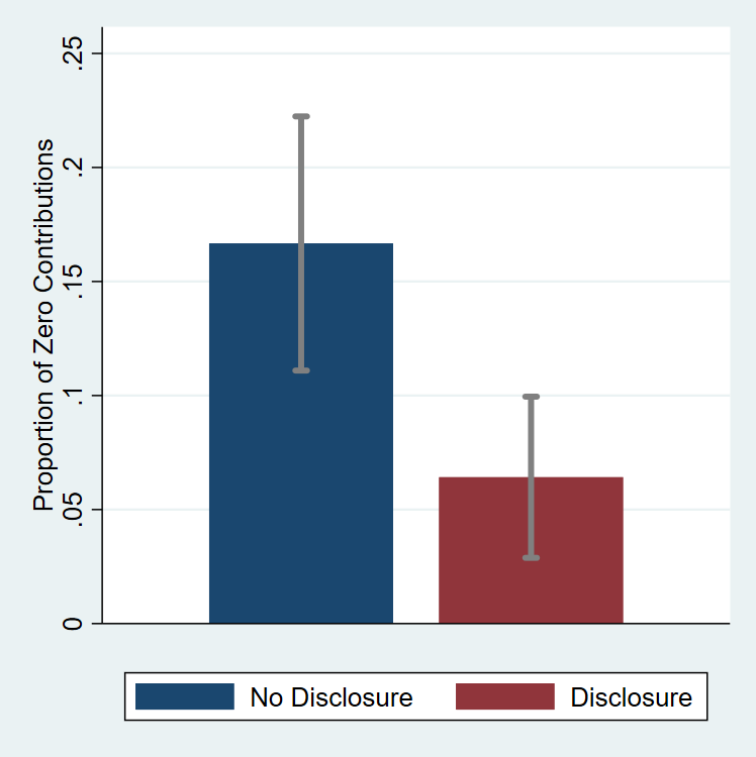


Figure 1: Proportion of free–riding (zero contribution) by ex–post disclosure (with 95% confidence intervals)

This provides evidence for "shame aversion" with aggregate behaviour being more cooperative (non–zero contributions) when being observed ex–post. A shame averse

person being one who experiences the negative emotion of shame if they do not meet the expectations of others (as in guilt aversion), but only when their behaviour is observed by others. By contributing more (through contributing a non–zero amount) the potential to feel shame by not meeting the expectation of one's partner is therefore reduced. This effect being in addition to any guilt aversion that would be present in either disclosure setting.[6]

Finding 2: *Average contributions do not significantly differ by ex–post disclosure.*

Support: Figure 2 reports the mean contribution level (out of 20 tokens) by ex–post disclosure. This includes all subjects so therefore it includes those who contribute nothing. It shows very little difference in aggregate and this is not significant at the 5% level when using a two–sided non–parametric Mann–Whitney–Wilcoxon (MWW) test ($p = 0.679$).

These average levels of contributions close to 50% are very similar to those typically seen in one–shot public goods games or the first round of multiple round games (Chaudhuri, 2011).

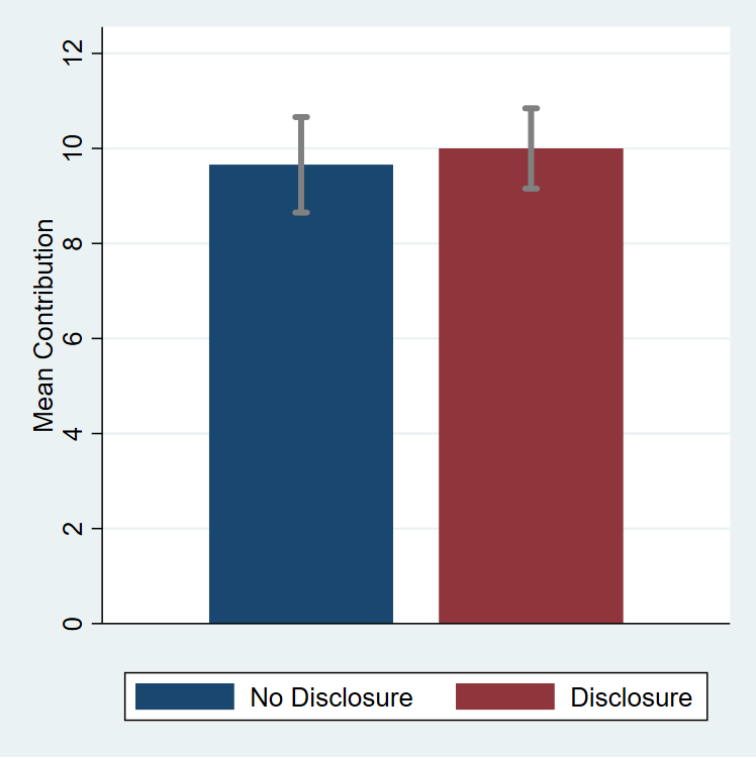


Figure 2: Mean contribution by ex–post disclosure (with 95% confidence intervals)

[6]Guilt not being dependent on being observed, as it has been defined throughout this paper

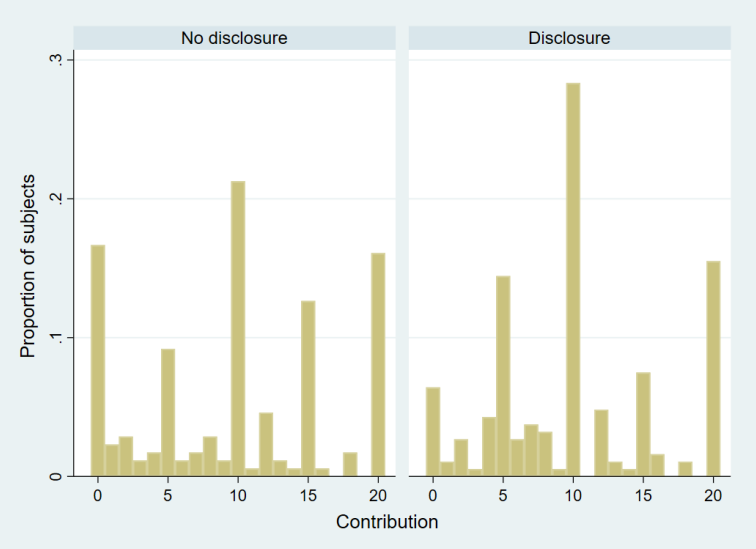


Figure 3: Histograms showing the distribution of contributions by ex–post disclosure

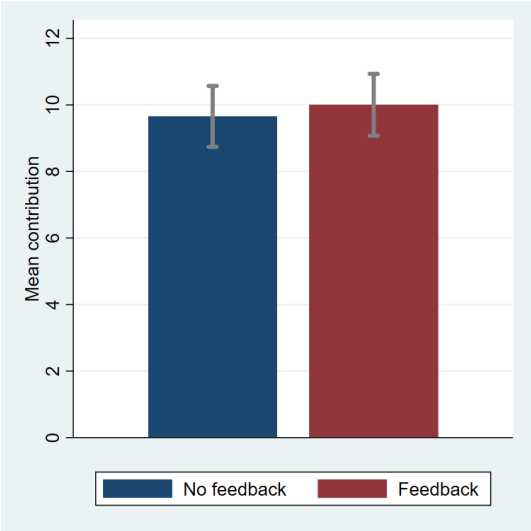
Contributions do also not differ significantly when looking solely at contributions above zero. This again being tested using a two–sided non–parametric Mann–Whitney–Wilcoxon (MWW) test ($p = 0.120$).

Figure 3 shows the distribution of contributions by ex–post disclosure and shows that the main affect related to ex–post disclosure is on those who would have made zero contributions (complete free–riding) with no ex–post disclosure and instead contribute a relatively small non–zero amount under disclosure. The effect on those that would make a non–zero contribution under either scenario appears to be minimal.

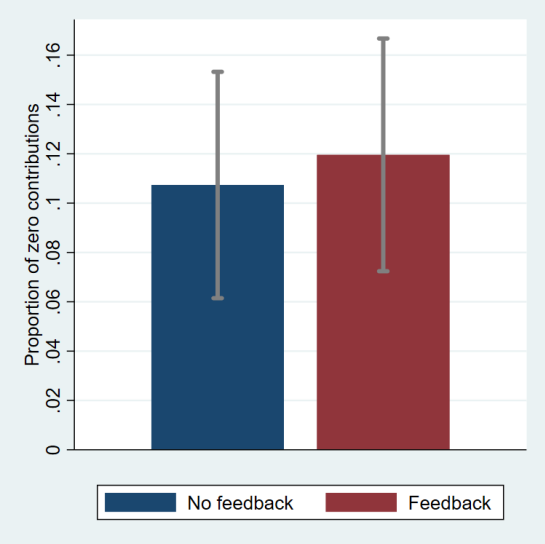
Finding 3: *Average contributions and contributions of zero do not significantly differ by ex–post feedback.*

Support: Figure 4a reports the mean contribution level by ex–post feedback where a subject knows that they will receive information on what their partner contributed. As a reminder, such ex–post feedback is present in the AD–I / FD roles and there is no ex–post feedback in the ND / AD–U roles. There is very little difference in aggregate and this is not significant at the 5% level when using a two–sided non–parametric Mann–Whitney–Wilcoxon (MWW) test ($p = 0.460$).

Figure 4b reports the proportion of complete free–riding (zero contribution) by ex–post feedback. There is again very little difference and this is not significant at the 5% level when using a two–sided Fisher's exact test ($p = 0.743$).



(a) Mean contribution



(b) Proportion of free–riding (zero contribution)

Figure 4: Mean contributions and free–riding by ex–post feedback (with 95% confidence intervals)

The tests used to support Findings 1, 2 and 3 pool together the different subject roles from across the experiment by either ex–post disclosure or ex–post feedback, i.e. it pools over the low and high social norm manipulations. As a more robust check and to explore if there are any interaction effects, a regression analysis is conducted in section 3.3 that supports the above findings.

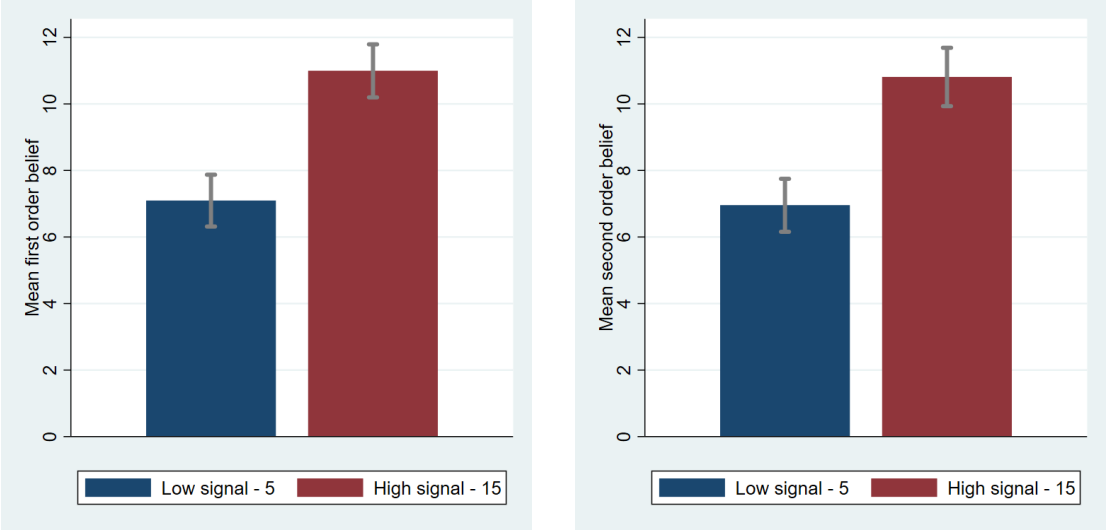
3.2 Social norm manipulation

Finding 4: *First and second order beliefs are on average higher under the high exogenous norm.*

Support: Figure 5 reports the mean first order (FOB) and second order (SOB) beliefs relating to the contribution level of their partner by the exogenous norm signal. It shows significantly higher beliefs under the high signal with the differences being significant at the 1% level when using two–sided non–parametric Mann–Whitney–Wilcoxon (MWW) tests (FOB: $p < 0.00001$, SOB: $p < 0.00001$). This shows that the exogenous norm treatment worked as intended and can therefore be used to examine the effect of beliefs on contributions in a way that is immune to any consensus effect concerns.

It is worth noting that the average beliefs in the low norm treatment are higher

than the actual norm given of five and the opposite is true for the high treatment, with beliefs below that of the norm of 15. This suggest that the low/high norm signal is shifting beliefs down/up respectively (on aggregate) but not to the full extent of beliefs matching the norm.



(a) First order beliefs (b) Second order beliefs

Figure 5: Mean first and second order beliefs by norm signal (with 95% confidence intervals)

Finding 5 (guilt aversion): *The higher norm signal significantly increases average contributions.*

Support: Figure 6 reports the mean contribution level (out of 20 tokens) by the exogenous norm signal. This signal being found in Finding 4 to lead to higher first and second order beliefs under the high norm signal. Figure 6 shows a significant difference with this being significant at the 1% level when using a two–sided non–parametric Mann–Whitney–Wilcoxon (MWW) test ($p < 0.00001$).

The same test conducted between the low and high signal treatments for each of the four roles is always significant at (at least) the 5% level ($p < 0.05$). This providing evidence that the norm signal has an effect on behaviour, even when subjects' choices were not observed by their partner (as in the ND and AD–I roles).

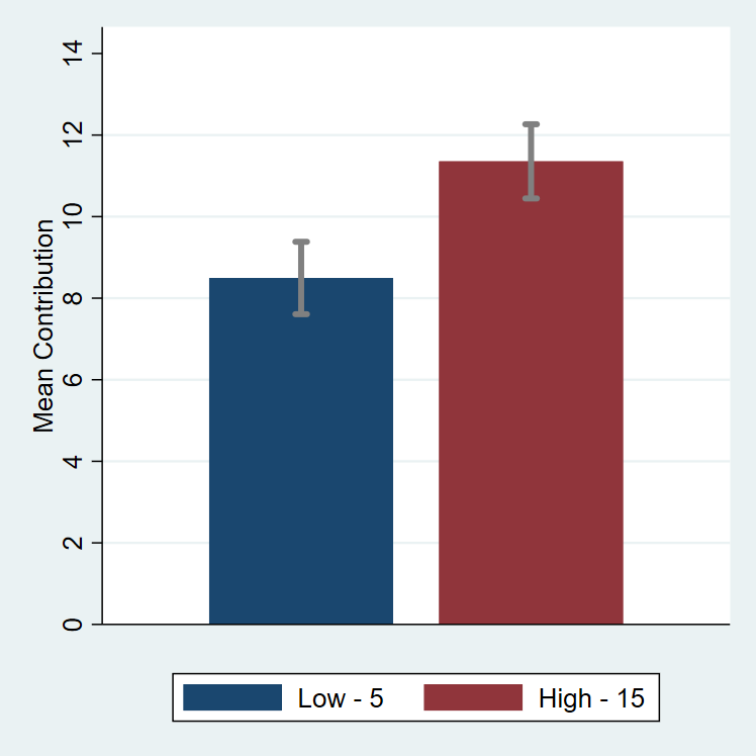


Figure 6: Mean contribution by norm signal (with 95% confidence intervals)

This result combined with Finding 4 supports the guilt aversion hypothesis that contributions are positively correlated with beliefs, at least on aggregate. This being tested in a way that is not confounded by any possible consensus effects as outlined previously.

3.3 Regression analysis

As the dependent variable of the number tokens contributed is bounded at zero and 20, a three–equation Cragg hurdle regression is used to model separately the choice to contribute more than zero (one hurdle), contribute below 20 (the second hurdle) and, lastly, unbounded contributions between and including one and 19 that are conditional on clearing the two hurdles at the bounds.

Table 2 reports the results of such a regression with dummy variables for ex–post feedback, ex–post disclosure and the high norm signal. The baseline being no ex–post feedback or disclosure and being in the low signal manipulation. Interaction variables are included for the high signal with feedback/disclosure.

Table 2: Bounded / unbounded contributions and treatments

	Participation: Positive contribution	Contribution (non–bounded)	Participation: Contribution below 20
Feedback	–0.153	–0.857	0.071
	(0.240)	(0.900)	(0.219)
Disclosure	0.668***	–0.757	0.145
	(0.243)	(0.760)	(0.220)
High	0.202	3.20***	0.269
	(0.288)	(0.950)	(0.280)
High × Feedback	0.207	2.62**	–0.145
	(0.361)	(1.17)	(0.320)
High × Disclosure	–0.257	–0.825	–0.360
	(0.368)	(1.07)	0.321)
N	361	263	361

Linear Cragg hurdle regression.
All independent variables used to model the outcome variance
Note: Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The leftmost column looks at the effect on the choice to contribute an amount greater than zero instead of zero. it can be seen that only ex–post disclosure is found to have a significant effect and this goes in the direction of reducing contributions of zero. This supports the earlier results, namely Findings 1 and 3.

Finding 5 is supported by the centre column looking at unbounded contributions which shows that the high norm signal does lead to significantly higher unbounded contributions.

Also significant is a positive interaction effect between the high signal and ex–post feedback suggesting that the effect of the high signal is reinforced when knowing that you will observe what the other subject that you are matched with chose. This is however the only evidence of any effect related to ex–post feedback.

The rightmost column reveals that there is no evidence of any significant effects on moving away from the upper bound of contributing the full amount of 20 tokens.

There is no clear evidence of any significant effects related to demographic characteristics such as age, gender and nationality.

4 Conclusions

This study has shown that there is evidence for exposure effects in the rather general social dilemma of a one–shot two–player public goods game. Specifically, "shame aversion" when people know they will be observed ex–post by those impacted by their decision. This works through less free–riding (contributions of zero) in the public goods game when there is such exposure. Contributions at an aggregate level are however relatively unaffected.

This occurs in a relatively low stakes, online 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.

This "shame aversion" builds on the hypothesis of guilt aversion. This predicts a positive relationship between contributions and second order beliefs about what an individual thinks others expect them to do. In line with most existing literature, robust evidence is found for this guilt aversion hypothesis. In contrast to earlier research however, this is found thorough a between–subject analysis that exogenously manipulates beliefs at a treatment level. This eliminating the potential for any confounding consensus effects or the need to use a potentially problematic "induced beliefs" methodology.

5 Appendix A. Experiment Instructions (Two separate screens: "Introduction" and "Instructions". See bold text to differentiate treatments/roles)

Introduction

Thank you for agreeing to take part in this study conducted in association with the University of Exeter which has received approval from the University of Exeter's research ethics committee.

The instructions presented here accurately describe the study. Please note that no personally identifiable information is collected and everyone will remain anonymous throughout and after this study.

We will process personal data in accordance with the EU General Data Protection Regulation (GDPR). By continuing, you consent to the publication of study results but that no data will be personally identifiable. Furthermore, anonymised data may be stored for an indefinite period of time and/or made available online to other researchers. Any published data will have your Prolific ID removed. This ensures anonymity and that data collected during the study cannot be linked to data from other studies run through Prolific. This does however also mean that you will not be able to withdraw consent after completing the study as there will be no way to link you to your data. You are however free to withdraw at any time during the study without giving a reason. To do so simply return your submission on Prolific and/or close the associated internet browser screens.

By completing the study you will earn at least £0.80 for participating. In addition, you may also earn more money, as described in the following instruction pages. Any additional earnings will be paid as a bonus payment.

The study must be completed in one go so please only continue if you have at least 10 minutes to complete it (however, it will most likely take less than this for many people). For the smooth running of the study, you must make your choices within a time limit on each page. The time remaining will be shown at the top of the page. If you do not make a decision before time runs out, then you will be redirected to a timeout screen and you will not be able to complete the study. If you do not

have time to complete the study or you do not consent to taking part, please return your submission now before continuing. By entering your Prolific ID and clicking next you agree that you understand the above information and give your consent to taking part in the study in accordance with the above notices.

Instructions

In these instructions your earnings are referred to in terms of Experimental Currency Units (ECUs). At the end of the experiment the total amount of ECUs that you have earned will be converted to Pounds at the rate, 1 ECU = £0.05.

With 50% probability you will be paid based on the choices made by you and another participant as described below. With 50% probability you will be paid a random amount between and including £0.80 and £1.80 to the nearest £0.01 with each amount being equally likely.

Once you click through to the next screen you will be randomly matched with another participant in the study. This other participant receives these exact same instructions.

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)

(ND treatment) Neither you nor the other participant will find out what choices or predictions the other made and neither of you will find out if you were paid based on the choices or randomly (see above).

(AD treatment) The contribution of one participant will be revealed to the other participant and one participant's contribution will never be revealed to the other participant. The participant receiving the other participant's contribution information will also find out if they were paid based on the choices or randomly (see above). The other participant does not find this out. **(AD-U subject)** You are the participant that will not receive any information relating to the other participant's contribution. Your contribution will however be revealed to the other participant and they will also find out if they were paid based on the choices or randomly. **(AD-I subject)** You are the participant that will receive information relating to the other participant's contribution. Your contribution will however never be revealed to the other participant.

(FD treatment) Both of you will be informed of each other's contributions and both of you will also find out if you were paid based on the choices or randomly (see above).

At the top of the screen where you make your decision, both you and the other participant will be told the average contribution choice made by participants in a similar experiment with the same payment structure. This average is the same for both of you.

You are also asked for two predictions. The first prediction is what you think the contribution of the other participant will be.

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

You will receive a bonus of 2 ECUs for each correct prediction and this will be paid whether or not you are paid randomly or not for the main part of the experiment.

The experiment will conclude with a short questionnaire.

Note that a copy of these instructions will be available on the next screen where you make your choices.

6 Appendix B. Control questions and answers (correct in bold)

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

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

3. Will both you and the other participant see the same average contribution choice of participants in a similar experiment? **Yes**, No.

4. Will the other participant ever find out what you contributed? ND / AD–U: Yes, **No**. AD–I / FD: **Yes**, No.

References

Andrighetto, G., Grieco, D., & Tummolini, L. (2015). Perceived legitimacy of normative expectations motivates compliance with social norms when nobody is watching. *Frontiers in psychology*, *6*, 1413.

Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, *97*(2), 170–176. doi: 10.1257/aer.97.2.170

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

Bracht, J., & Regner, T. (2013). Moral emotions and partnership. *Journal of Economic Psychology*, *39*, 313–326.

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

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental economics*, *14*(1), 47–83.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

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.

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*.

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.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of experimental social psychology*, *13*(3), 279–301.

Staudt, A. (2019). CIBAR: Stata module to plot bar graphs and confidence intervals over groups.

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

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