Dhami, Sanjit; Wei, Mengxing; al-Nowaihi, Ali

Working Paper

Public Goods Games and Psychological Utility: Theory and Evidence

CESifo Working Paper, No. 7014

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Dhami, Sanjit; Wei, Mengxing; al-Nowaihi, Ali (2018) : Public Goods Games and Psychological Utility: Theory and Evidence, CESifo Working Paper, No. 7014, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/180276

CESifo WORKING PAPERS

7014
2018

April 2018

Public Goods Games and Psychological Utility: Theory and Evidence

Sanjit Dhami, Mengxing Wei, Ali al-Nowaihi

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7014*
Category 13: Behavioural Economics

Public Goods Games and Psychological Utility: Theory and Evidence

Abstract

We consider a theoretical model of a public goods game that incorporates reciprocity, guilt-aversion/surprise-seeking, and the attribution of intentions behind these emotions. In order to test our predictions, we implement the 'induced beliefs method' and a within-subjects design, using the strategy method. We find that all our psychological variables contribute towards the explanation of contributions. Guilt-aversion is pervasive at the individual-level and the aggregate-level and it is relatively more important than surprise-seeking. Our between-subjects analysis confirms the results of the within-subjects design.

JEL-Codes: D010, D030, H410.

Keywords: public goods games, psychological game theory, reciprocity, surprise-seeking/guilt-aversion, attribution of intentions, induced beliefs method, within and between subjects designs.

*Sanjit Dhami**
Division of Economics, School of Business
University of Leicester
University Road
United Kingdom – Leicester LE1 7RH
sd106@le.ac.uk

Mengxing Wei
International Academy of Business and Economics, Tianjin University of Finance and Economics, 25 Zhujiang Road
P.R. of China – Tianjin, 300222
maxine.wmx@gmail.com

Ali al-Nowaihi
Division of Economics, School of Business
University of Leicester
University Road
United Kingdom – Leicester LE1 7RH
aa10@le.ac.uk

*corresponding author

January 2018
This is the working paper version of a paper that is forthcoming in the Journal of Economic Behavior and Organization in a special issue on Psychological Game Theory. We are very grateful to two perceptive referees and the editors, Amrish Puri and Martin Dufwenberg. We would also like to acknowledge several helpful comments and suggestions during talks given at Ludwig Maximilian University of Munich, Innsbruck University, Tilburg University, and the 2nd Psychological Game Theory Workshop in UEA, Norwich.

1. Introduction

In classical game theory, the utility functions of the players map the set of strategy profiles into the set of payoffs. Since beliefs do not directly enter the utility function, we shall refer to utility in classical game theory as *material utility*. In contrast, a range of phenomena are more satisfactorily explained by introducing beliefs directly into the utility function of players. Beliefs are important in classical game theory too. For instance, Bayesian updating is used to update beliefs along the path of play, but beliefs do not directly enter into utility functions. The following example illustrate how the feelings of *surprise* and *guilt* may directly impart disutility.

Example 1 *: John frequently visits cities A and B, and he typically uses a taxi to get around. In city A, tipping a taxi driver is considered insulting, while in city B it is the norm to tip a publicly known percentage of the fare. Suppose that it is common knowledge that if taxi drivers do not receive a tip, they quietly drive away. In city A, John gives no tip, and feels no remorse from not giving it. However, in city B, the taxi driver expects John to give him a tip (taxi driver's first order belief) and John believes that the taxi driver expects a tip from him (John's second order belief). Based on his second order belief, John cannot bear the guilt of letting the taxi driver down by not paying the tip. Thus, he tips every time he takes a taxi in city B. Clearly, John's utility appears to be directly influenced by his second order beliefs.*

Players may also derive psychological utility or disutility from a range of other emotions relating to kindness, anger, surprise, malice, joy, and hope, that can be captured by defining appropriate beliefs in the game (Elster, 1998). Our main focus in this paper shall be on *reciprocity*, *guilt–aversion/surprise–seeking* and on the *attribution of intentions* behind these emotions. We formally define these concepts below.

The proper theoretical framework to deal with these issues is *psychological game theory* (Geanakoplos et al., 1989; Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Battigalli and Dufwenberg, 2009). This is not simply a matter of augmenting material payoffs with beliefs of various orders and then applying the classical machinery in game theory. This is because beliefs themselves are endogenous, hence, an entirely new framework is needed.

Reciprocity was developed for simultaneous move games by Rabin (1993) and extended to sequential games by Dufwenberg and Kirchsteiger (2004). In Example 1, in city B, if John believes that the taxi driver has been particularly courteous and helpful, then he might tip him extra, on account of *reciprocity*.

Battigalli and Dufwenberg (2007) proposed a formal approach to modelling guilt. In particular, they highlight two different emotions associated with guilt.

(1) *Simple guilt* arises from falling short of the perceived expectations of other players. For instance, if in city B in Example 1, John believes that the taxi driver expects a 15% tip, yet pays only a 10% tip, then he may suffer from simple guilt, which directly reduces his utility.

(2) *Guilt from blame* arises when one cares for the attribution of intentions behind psychological feelings such as guilt–aversion/surprise–seeking. In terms of Example 1, suppose that it is not common knowledge that taxi drivers who fail to receive a tip, drive away quietly. Instead, suppose that there is the possibility that John gives a tip purely because he prefers not to have an unpleasant argument with the taxi driver over a tip. In this case, the taxi driver must factor in the intentionality behind John's psychological feelings, such as guilt, in giving the tip. In turn, John may derive direct disutility if he believes that his tip was believed by the taxi driver to be unintentional in the sense that it was given to avoid a potential argument. However, this requires the use of third order beliefs of the taxi driver and John's fourth order beliefs. Higher order beliefs require relatively greater cognitive resources on the part of players. Whether players use such higher order beliefs is an empirical question.

The surprise–seeking motive was formally identified by Khalmetski et al. (2015) in dictator game experiments. They also provide a theoretical framework in which surprise–seeking may be analyzed. The surprise–seeking motive arises from exceeding the expectations of others, as perceived by a player through his/her second order beliefs. For instance, in Example 1, in city B, John may believe that the taxi driver expects a tip that is 10% of the fare, yet he derives extra utility by offering instead a 15% tip (surprise–seeking motive). One may extend these beliefs to higher orders by factoring in the intentionality of surprise–seeking motive.[1]

Central to empirically identifying the guilt–aversion and/or the surprise–seeking motives is to specify the method of eliciting the beliefs of players. The simplest way of eliciting beliefs is to directly ask players their beliefs. This is the *self-reporting method* or *the direct elicitation method*. Empirical studies using the self-reporting method have given strong support for the simple guilt–aversion motive in various versions of the trust game, as well as in public goods games. Denote by ρ the correlation coefficient between one's actions and one's second order beliefs (i.e., beliefs about the other player's first order beliefs). The typical finding that supports guilt–aversion is a statistically significant and *positive* value of ρ.[2]

[1] For a treatment of psychological game theory and more examples, see Chapter 13 in Dhami (2016).

[2] For trust game experiments that support this finding, see Dufwenberg and Gneezy (2000), Guerra and Zizzo (2004) and Reuben et al. (2009). For supporting evidence from public goods games experiments, see Dufwenberg et al. (2011).

Ellingsen et al. (2010) pose an important challenge to models of guilt–aversion by questioning the validity of the self-reporting method. They argue that self-reported second order beliefs of players, i.e., beliefs about the first order beliefs of others are subject to the *false consensus effect* (Ross et al., 1977). This is also known as *evidential reasoning* and the relevant theory is formalized in al-Nowaihi and Dhami (2015). The argument is that people take their own actions as diagnostic evidence of what other like-minded people are likely to do. In terms of Example 1, John is subject to the false consensus effect if in forming his second order beliefs about the tip expected by the taxi driver, he assigns his own propensity to tip the taxi driver as the relevant first order beliefs of the taxi driver. Indeed there might be no relation between the taxi driver's actual first order beliefs and John's propensity to give a tip to the taxi driver.

In order to support their argument, Ellingsen et al. (2010) implement a radical experimental design. In the first stage, they directly ask players for their first order beliefs about the actions of the other player in two-player games (dictator, trust and a partnership games). These beliefs are then revealed to the other player before they make their decision. Players are given no information about how their beliefs will be used, so it is hoped that beliefs are not misstated to gain a strategic advantage. Thus, the second order beliefs of players (beliefs about the first order beliefs of others) are as accurate as possible. It is as if players can peep into the minds of other players to accurately gauge their beliefs.[3] One might wonder if this experimental design constitutes subject deception; Ellingsen et al. (2010) give a robust defence of their procedure against such a charge.[4] We term this method of belief elicitation as the *induced beliefs method* in comparison with the earlier self-reporting method.

Ellingsen et al. (2010) then showed, using the induced beliefs method, that the correlation between second order beliefs and actions, ρ, is not statistically different from zero. They draw the following two conclusions. (i) Guilt-aversion is absent. (ii) Earlier studies on guilt–aversion that use the self-reported beliefs method may just have been picking the false consensus effect, which traditionally lies outside psychological game theory. These findings were, at that time, a devastating critique of the ability of psychological game theory to explain economic phenomena, at least those that involved guilt–aversion.

[3]This design is not subject to other confounding influences. For instance, pre-play communication may enhance first and second order beliefs (Charness and Dufwenberg, 2006). Yet pre-play communication might influence actions not because players suffer from guilt-aversion, but rather because they may have a preference for promise-keeping (Vanberg, 2008).

[4]Technically, subjects were not lied to. They were simply not given any information about how their beliefs would be used. In the exit interviews, none of the subjects complained about being misled, once they were told that their first order beliefs were revealed to the other player. There could, however, be externalities for other experimenters if the same subjects participate in other experiments. The authors assign little probability to such an event.

Khalmetski et al. (2015) stick with the induced beliefs method and the dictator game (both used in Ellingsen et al., 2010). They argued, and showed, that the Ellingsen et al. (2010) findings can be reconciled with models of psychological game theory if we also recognize, in addition, the surprise–seeking motive. The main testable implication in this case is derived by eliciting the transfers made by dictators as they receive different signals of the first order beliefs of the passive receivers. Since the predictions of the model are for the behavior of individual dictators, a *within–subjects design* was implemented with the *strategy-method* (in contrast, Ellingsen et al., 2010, used a between–subjects design). For their overall sample, they find that ρ is not significantly different from zero (as in Ellingsen et al., 2010), but the situation is different at the individual level. When psychological factors are statistically significant, about 70% of the dictators are guilt-averse and about 30% are surprise–seeking. However, the behavior of the two types of players cancels out in the aggregate, giving rise to the appearance that ρ is not significantly different from zero. Thus, the Ellingsen et al. (2010) results were shown to be too aggregative to pick out individual level guilt aversion.[5]

The existing literature, and the state of the art, as described above, have the following two main features that motivate our paper.

1. *Portability of the dictator game results:* Ellingsen et al. (2010) and Khalmetski et al. (2015) use dictator game experiments in which there is no element of strategic interaction.[6] In justifying the use of the dictator game for their problem as a useful starting point, Khalmetski et al. (2015, p. 166) write: "... it abstracts away from potentially confounding strategic or reciprocal interaction." However, we know from many contexts that the results from dictator games may lack robustness and may not survive the introduction of strategic elements. Despite its popularity, the dictator game might not be a particularly good game to test alternative theories that require even a modicum of strategic interaction (Fehr and Schmidt, 2006; Dhami, 2016, Part 2).

2. *Difficulty of comparing different methodologies:* In studies that use the induced beliefs method, there is a lack of uniformity among studies about the methodology with respect to the within–subjects or between–subjects design. Khalmetski et al. (2015) use a within–subjects design in testing for simple guilt–aversion/surprise–

[5] In a recent paper, Khalmetski (2016) proposes another method of inferring guilt aversion in two-player sender-receiver games where one player has perfect information about the game, but the other player's imperfect information is varied by providing selective information on the parameters of the game. This, in turn, induces an exogenous variation in the second order beliefs of the player, which can be correlated with actions to infer guilt-aversion.

[6] Ellingsen et al. (2010) also report results from a trust game but only the dictator game results are comparable with Khalmetski et al. (2015).

seeking but a between–subjects design for the role of attributions behind intentions about guilt–aversion/surprise–seeking. In contrast, Ellingsen et al. (2010), who did not test for the role of attribution of intentions behind guilt, use a between–subjects design throughout.

In light of the two features discussed above, two natural questions, that lie at the heart of our paper, are as follows.

1. Do the theoretical and empirical results of Ellingsen et al. (2010) and Khalmetski et al. (2015) extend to games with strategic components, such as public goods games? In addition to this, we are also interested in modelling reciprocity within an encompassing framework that includes guilt–aversion and surprise–seeking.[7]

2. If each of the two psychological components, simple guilt–aversion/surprise–seeking and attribution of intentions behind guilt–aversion/surprise–seeking, are tested in a within–subjects and a between–subjects design, then how do the results compare? Does this help us to reconcile conflicting experimental findings?

We address the first question by considering a public goods game, which has an explicit strategic interaction component. We first extend the theoretical framework of Khalmetski et al. (2015), designed for dictator games, to a two-player public goods game.[8] We also consider issues of reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004). Our framework extends readily to many players, but we prefer the two-player game for the following reasons. First, the existing empirical results come from two-player games such as dictator and trust games. Second, for public goods games with three players or more, players need to form beliefs about the beliefs of other players about all the opponents, which is cognitively more challenging. Hence, we believe that our model and empirical tests provide a cleaner, sharper test of the relevant theory and a better comparison with the existing literature.

We address the second question by considering a within–subjects design and a between–subjects design for each of our main treatments. This allows us to give a more satisfactory account of the predictions of psychological game theory for public goods games and also facilitates comparison with the existing literature.

In an induced beliefs design, the within–subjects findings from our public goods game experiments are as follows. Reciprocity, as reflected in the effect of first order beliefs on

[7]There is of course a large classical literature on public goods games in which beliefs do not play a central role (Dhami, 2016, Sections 5.4, 5.5). But we are mainly interested in the role of beliefs in the framework of psychological game theory.

[8]We use the same additive belief structure as in Khalmetski et al. (2015). Hence, many of their important results, particularly the contrast between the private and public treatments, carry over in a natural fashion to our analysis.

contributions, is a significant motivation. We also find that overall, across all subjects, actions are significantly influenced by second order beliefs, thus, there is also guilt–aversion in the aggregate. Furthermore, we find that for the statistically significant cases, for 95% of our subjects, guilt–aversion is relatively more important, but for the remaining 5% subjects the surprise–seeking motive is relatively more important. Thus, in our data, guilt–aversion is, by far. the predominant finding at the level of the individuals and for the aggregate data. These findings are in contrast with the findings of Ellingsen et al. (2010) and Khalmetski et al. (2015) for dictator games that we report above; in particular, both studies find no significant aggregate guilt–aversion.

We find that, for at least 30% of our subjects, the *attribution of intentions* behind guilt–aversion/surprise–seeking is statistically significant, although we cannot rule out this motive for our remaining subjects.

Our between–subjects design with induced beliefs replicates the results of our within–subjects design. A regression analysis shows that first order beliefs and second order beliefs significantly and positively influence contributions; this is consistent with, respectively, reciprocity and guilt–aversion. In contrast, for the dictator game with induced beliefs and a between–subjects design, Ellingsen et al. (2010) find the absence of guilt–aversion. Khalmetski et al. (2015) also replicated these findings of no aggregate guilt–aversion for dictator games in a similar design although they report important within-subject heterogeneity.

In an induced beliefs design, Khalmetski et al. (2015) found that when the first order beliefs of the passive receiver are publicly announced in a public treatment, then, in aggregate, dictators make larger transfers relative to a private treatment where such beliefs are not publicly announced. In contrast, we find, in our public goods game, that in both our within–subjects and between–subjects designs, there is no statistically significant aggregate difference in public goods contributions relative to the two private treatments. However, at the individual level there is heterogeneity among individuals in the extent of contributions if the beliefs are publicly announced.

We also explore some of the other determinants of contributions to public goods. There are significant gender effects; men contribute significantly less than women in the public treatment although there are no such differences in the private treatment. However, other factors are not significant in determining contributions. These factors include educational attainment, degree program in the university, and previous experience of participating in similar experiments.

The plan of the paper is as follows. Section 2 outlines the basic model of public goods. Section 3 considers the implications of reciprocity, guilt–aversion, and surprise–seeking in a two-player public goods game; we also consider the psychological motivation of the attribution of intentions behind guilt–aversion, and surprise–seeking. Section 4 gives the theoretical predictions of our model. Section 5 describes our within–subjects

experimental design and Section 6 gives our experimental results. Section 7 reconsiders the empirical results in a between–subjects design. Section 7.4 examines the determinants of contributions. In Section 8, we attempt to reconcile our results on the proportion of subjects who exhibit guilt-aversion/surprise-seeking with the results from previous studies. This opens up a potentially novel channel through which signals about the first order beliefs of others may enhance one's own prosocial behavior. Section 9 concludes. Appendix A contains the proofs. Appendices B and C contain, respectively, the instructions for the within–subjects and the between–subjects designs. Appendix D provides further discussion of the psychological utility functions.

2. The classical model of public goods

Consider a public goods game with two players $N = \{1, 2\}$. We use the index $i = 1, 2$ for the players. Variables pertaining to player i are subscripted by i and variables pertaining to the other player by $-i$. Each player has an initial endowment of $y > 0$ monetary units.[9] The two players simultaneously choose to make contributions $g_i \in [0, y], i = 1, 2$, towards a public good. The production technology is assumed to be linear, so the total production of the public good is $G = g_1 + g_2$. The utility function is quasilinear and given by $u_i : [0, y]^2 \to \mathbb{R}$. In particular, $u_i(g_i, g_{-i}) = v_i(c_i) + r(g_i + g_{-i})$, where $r > 0$, and v_i is a strictly increasing and strictly concave utility function of private consumption, c_i, so $v_i' > 0$, $v_i'' < 0$. The budget constraint is given by $c_i + g_i = y$. Substituting the constraint into the utility function, the utility function becomes

$$u_i(g_i, g_{-i}) = v_i(y - g_i) + r(g_i + g_{-i}). \tag{2.1}$$

The parameter r is interpreted as the unit return from the public good to each player; this captures the non-rival and non-excludable nature of the public good. We assume that $r < v_i'(y)$, i.e., the net return to an individual from a unit of contributions is negative. Since $v_i'' < 0$, thus,

$$0 < r < v_i'(y - g_i) \text{ for all } g_i \in [0, y]. \tag{2.2}$$

We state the benchmark result under the classical model, below, using superscript n on variables to denote the Nash equilibrium solution.

Proposition 1 *: In a Nash equilibrium of the simultaneous move public goods game, each player chooses to free-ride and make a zero contribution, so* $(g_1^n, g_2^n) = (0, 0)$*, and total public good provision is* $G^n = 0$*.*

To distinguish the ordinary utility (2.1) from the psychological utility, to be introduced in Section 3, immediately below, we shall refer to (2.1) as the *material utility*.

[9] In our experiments, the endowment is expressed in tokens. All subjects are made aware of the exchange rate between tokens and money.

3. The model of public good contributions under surprise–seeking and guilt–aversion

In this section we introduce the assumptions behind our model of public good contributions in the presence of psychological tendencies such as surprise–seeking and guilt–aversion.

3.1. Levels of beliefs

We now modify the classical model to incorporate the emotions that arise from the positive surprises (surprise–seeking) and the negative surprises (guilt–aversion) that players cause for others, relative to a reference point that we describe below. The beliefs of each player are private information to the player, although players may (and in our model some do) observe signals of other's beliefs. The basic structure of beliefs is similar to that in Khalmetski et al. (2015).

The beliefs are defined recursively as follows.

I. First order beliefs: Let b_i^1 be the first order belief of player $i = 1, 2$ about the level of contribution, g_{-i}, of the other player. The cumulative distribution of b_i^1 is $F_i^1 : [0, y] \to [0, 1]$.

II. Second order beliefs: Let b_i^2 be the second order belief of player $i = 1, 2$ about the first order belief of the other player, b_{-i}^1. The cumulative distribution of b_i^2 is $F_i^2 : [0, y] \to [0, 1]$. However, before forming second order beliefs, player $i = 1, 2$ may observe a signal θ_i of the first order belief distribution of the other player, F_{-i}^1. Since players may alter their beliefs based on the signal that they receive, we are also interested in their conditional beliefs. Let $F_i^2(x|\theta_i)$ be the conditional cumulative distribution of the second order belief, b_i^2, of player i about the first order belief, b_{-i}^1, of the other player.[10].

III. Third and fourth order beliefs: Let b_{-i}^3 be the third order belief of player $-i$, $i = 1, 2$, about the second order belief, b_i^2, of player i. The cumulative distribution of b_{-i}^3 is $F_{-i}^3 : [0, y] \to [0, 1]$. Ex-post, player $-i$ observes the contributions, g_i, made by player i and must infer the intentionality behind this choice, which requires the use of $F_{-i}^3(x)$. However, player i does not know $F_{-i}^3(x)$ when choosing g_i, hence, he uses his beliefs about $F_{-i}^3(x)$, given by $F_i^4(x)$, in forming expectations about player $-i's$ beliefs about his intentions.[11] In subsection 3.2.2, below, we shall introduce conditional fourth order beliefs.

[10] Specifically, $F_i^2(x \mid \theta_i)$ is the probability assigned by player i that the first order belief of the other player, b_{-i}^1, takes a value less than or equal to $x \in [0, y]$, conditional on θ_i.

[11] In principle, one may define beliefs up to any order (as in fact required in classical game theory). However, we need beliefs up to order 4 only because we are mainly interested in the emotions of guilt and the attributions of intentions behind guilt-aversion and surprise-seeking.

3.2. Treatments

We have three treatments: The *asymmetric private treatment* (APR), the *private treatment* (PR) and the *public treatment* (PUB). The treatment PR, that is related to the treatment APR, is used only in our between–subjects design, so we postpone a discussion of it to Section 7. We now discuss the other two treatments that are common to the within–subjects and the between–subjects design.

3.2.1. APR treatment

In APR, subjects are divided into two equal groups: APR1 and APR2. Every subject in APR1 is randomly matched, one to one, with a subject from APR2 to play the two-player public goods game. We shall use the subscript 1 to denote a player in APR1 (or player 1) and a subscript 2 to denote a player in APR2 (or player 2). Players in APR1 are the informed players. Player 1 receives a signal, $\theta_1 \in [0, y]$, about the contribution, $g_1 \in [0, y]$, that player 2, expects him to make. Player 2 does not know that player 1 has received this information. Furthermore, player 1 knows that player 2 does not know that player 1 has received this information. Player 2, by contrast, does not receive any signal about the expectation of player 1 about his (player $2's$) contribution.

Player 1 derives utility from believing that his actual contribution, g_1, is greater than what player 2 expected him to contribute (*simple* surprise–seeking). Player 1 also derives disutility from believing that g_1 is less than what player 2 expected him to contribute (guilt–aversion). For this, player 1 has to form a second order belief, $F_1^2(x)$, about what player 2 expects. Before choosing g_1, player 1 receives a signal, θ_1, about player 2's expectation of g_1. Hence, player 1 can update his belief by conditioning on this signal. So, the relevant distribution for him is the conditional distribution $F_1^2(x|\theta_1)$. Player 2 also experiences similar emotions of simple surprise–seeking and guilt–aversion; the only difference is that player 2 does not receive any signal from player 1.

Ex-post, after all contribution decisions have been made, the contribution, g_1, of player 1 is communicated to his partner, player 2. Player 1 derives utility from believing that player 2 thinks that player 1 intended to positively surprise him (*intentional* surprise–seeking). Player 1 also derives disutility from believing that player 2 thinks that player 1 intended to negatively surprise him (*intentional* guilt–aversion). For this, player 1 has to form a fourth order belief, $F_1^4(x)$, about the third order beliefs of player 2, $F_2^3(x)$ (i.e., what player 2 thinks player 1 believes player 2 expected him to contribute).

Notice that the relevant fourth order beliefs are the unconditional beliefs. The reason is that player 1 in the APR treatment knows that player 2 is unaware that player 2's guess is revealed as a signal θ_1 to player 1. Thus, the third order beliefs of player 2, $F_2^3(x)$ (which are beliefs about $F_1^2(x)$), must be independent of θ_1. This implies that F_1^4, which

are beliefs about F_2^3, must also be independent of θ_1.

3.2.2. PUB treatment

In contrast to the APR treatment, in the PUB treatment, each player, $i = 1, 2$, receives a signal, θ_i, about the contribution, g_i, that his partner, player $-i$, expects him (player i) to make. Furthermore, since the signals are publicly announced and players know that they are publicly announced, the signals are public knowledge. This implies that as compared to the APR treatment, in the PUB treatment, the relevant fourth order beliefs are the conditional beliefs $F_i^4(x|\theta_i)$. The reason is that public knowledge of the transmission of signals ensures that the third order belief of player $-i$, F_{-i}^3, depend on the signal θ_i. In turn, F_i^4, the belief of player i about the third order beliefs of player $-i$, must also depend on θ_i.

3.3. Assumptions on beliefs

We make the following assumptions.

Assumption A1 Beliefs are continuously distributed, i.e.,

$$f_i^k(x) \text{ is continuous on } [0, y], \; k = 1, 2, 3, 4, \; i = 1, 2.$$

Assumption A2 $F_i^2(x|\theta_i)$ and $F_i^4(x|\theta_i)$ are differentiable in θ_i, $i = 1, 2$.

Assumption A3 A higher value of the signal, θ_i, induces strict first order stochastic dominance in the conditional distribution of beliefs $F_i^2(x|\theta_i)$ and $F_i^4(x|\theta_i)$. Thus, we have

$$\theta_i' > \theta_i \Rightarrow F_i^k(x|\theta_i') < F_i^k(x|\theta_i) \text{ for all } x \in (0, y) \text{ and } \theta_i', \theta_i \in [0, y], \; i = 1, 2, \; k = 2, 4.$$

Since $F_i^k(x)$ is the integral of $f_i^k(x)$, it follows from the continuity of $f_i^k(x)$ that $F_i^k(x)$ is differentiable. However, it does not follow that $F_i^k(x|\theta_i)$ is differentiable in θ_i, which we shall need. Hence, we have explicitly stated this in Assumption A2.

Assumptions A2 and A3 imply that

$$\frac{\partial F_i^k(x|\theta_i)}{\partial \theta_i} < 0 \text{ for all } x \in (0, y) \text{ and all } \theta_i \in (0, y), \; i = 1, 2, \; k = 2, 4. \tag{3.1}$$

Except for Proposition 5, which reports the comparative statics with respect to θ_i, below, none of our results depend on how the first order belief, b_i^1, depends (if at all) on the signal θ_i. However, Proposition 5 depends crucially on the assumption that b_i^1 is independent of θ_i, i.e., $F_i^1(x|\theta_i) = F_i^1(x)$, $i = 1, 2$. We thus introduce Assumption 4, immediately below.

Assumption A4 : $F_i^1(x|\theta_i) = F_i^1(x)$, $i = 1, 2$.

In Section 8, we shall replace Assumption A4 by the Assumption A5, $\frac{\partial F_i^1(x|\theta_i)}{\partial \theta_i} < 0$, and derive Proposition 9, the appropriate reformulation of Proposition 5 under Assumption 5. This allows us to reconcile some conflicting empirical results in the literature on the relative incidence of surprise–seeking and guilt-aversion. With the exception of Propositions 5 and 9, none of our other results depend on whether we combine assumption A1-A3 with A4 or A5.

Khalmetski et al. (2015) assume that θ_i is the median of $F_i^2(x|\theta_i)$, i.e., $F_i^2(\theta_i|\theta_i) = \frac{1}{2}$. We do not need this assumption. In our formulation, θ_i could be any signal, such as the median, as in Khalmetski et al. (2015), or the average, or the mode (the most probable value) or any other statistic, provided Assumption A3 is satisfied.

Example 2 *: We consider a two-player public goods game. Each player has the initial endowment $y = 2$. Player i contributes $g_i \in [0, 2]$ to the public good, $i = 1, 2$. We consider the asymmetric private treatment (APR) where player 1 is the informed player (a member of APR1) and player 2 is the uninformed player (a member of APR2). Player 2 has a first order belief about the contribution, g_1, likely to be made by player 1 that is given by the probability density $f_2^1(x)$, $x \in [0, 2]$. This probability density is not known to player 1, who forms a second order belief about what player 2 expects player 1 to contribute. This second order belief of player 1 is given by the probability density $f_1^2(x)$, $x \in [0, 2]$; $f_1^2(x)$ may bear little similarity to $f_2^1(x)$.*

Player 2 makes a guess, $\theta_1 \in [0, 2]$, about the contribution player 1 will make. Unsure about what player 1 will contribute, player 2 reports a statistic about $f_2^1(x)$, for example the mean, the median, or the mode (or any other statistic) of his privately known belief distribution. Having received the signal θ_1, player 1 updates his belief by using the conditional distribution $f_1^2(x|\theta_1)$. In this Example, we shall assume that player 1 believes that θ_1 is what player 2 regards as the most probable value for g_1. Khalmetski et al. (2015) assume that θ_1 is the median of $f_1^2(x|\theta_1)$. But nothing in our paper depends on this assumption. For us, any statistic will do, provided that it satisfies Assumption A3. Moreover, player 1, being ignorant of the statistic chosen by player 2, may use a different statistic. For example, in this Example, player 1 assumes that player 2 reports the most probable value when, in fact, player 2 could have reported the median or average (or any other statistic). For the purposes of this Example, we take the second order belief of player 1 to have the conditional probability density:

$$f_1^2(x|\theta_1) = \frac{x}{\theta_1},\ x \in [0, \theta_1],\ \theta_1 \in (0, 2], \tag{3.2}$$

$$f_1^2(x|\theta_1) = \frac{2 - x}{2 - \theta_1},\ x \in [\theta_1, 2],\ \theta_1 \in [0, 2). \tag{3.3}$$

Geometrically, the density (3.2), (3.3) forms the two sides of a triangle with base length 2 *and height* 1 *(so the area under the density is* 1*, as it should be). The apex of the triangle is at* θ_1*. Hence, given that player* 1 *receives the signal* θ_1*, player* 1 *thinks that player* 2 *believes that the most probable value of player* 1*'s contribution,* g_1*, is* θ_1*.*
Suppose, for instance, that player 1 *receives* $\theta_1 = 2$*. In this case, player* 1 *thinks that player* 2 *believes that player* 1 *will most probably make the maximum contribution,* $g_1 = 2$*. From (3.2) we get* $f_1^2(x|2) = \frac{x}{2}$*,* $x \in [0, \theta_1]$*. At the other extreme, suppose player* 1 *receives* $\theta_1 = 0$*. Here, player* 1 *thinks that player* 2 *believes that player* 1 *will most probably contribute nothing,* $g_1 = 0$*. From (3.3) we get* $f_1^2(x|0) = 1 - \frac{x}{2}$*,* $x \in [\theta_1, 2]$*. The cumulative conditional distributions corresponding to (3.2) and (3.3) are, respectively,*

$$F_1^2(x|\theta_1) = \frac{x^2}{2\theta_1},\ x \in [0, \theta_1],\ \theta_1 \in (0, 2]. \tag{3.4}$$

$$F_1^2(x|\theta_1) = \frac{2x - \frac{1}{2}x^2 - \theta_1}{2 - \theta_1},\ x \in [\theta_1, 2],\ \theta_1 \in [0, 2). \tag{3.5}$$

From (3.4) and (3.5), it is straightforward to show:

$$\frac{\partial F_1^2(x|\theta_1)}{\partial \theta_1} < 0, x \in (0, 2), \theta_1 \in (0, 2), \tag{3.6}$$

in agreement with Assumption A3. Furthermore, by algebraic means, it is straightforward to show that any distribution, $F_1^2(x|\theta_1)$*, with* $\theta_1 \in (0, 2)$ *strictly first order dominates* $F_1^2(x|0)$ *and is strictly first order dominated by* $F_1^2(x|2)$*. Thus, Assumption A3 is satisfied.*
A large number (in fact, an infinite number) of unconditional distributions are consistent with (3.2)-(3.5). For example, let player 1*'s prior distribution of* θ_1 *(before he received the signal containing a realization of* θ_1*) be:*

$$\pi_1^2(\theta_1) = 1 - \frac{1}{2}\theta_1, \theta_1 \in [0, 2], \tag{3.7}$$

According to (3.7), player 1 *believes that player* 2 *thinks that the most probable contribution of player* 1 *is zero. But many other prior distributions are consistent with (3.2)-(3.5), including:*

$$\pi_1^2(\theta_1) = \frac{1}{2}\theta_1, \theta_1 \in [0, 2], \tag{3.8}$$

according to which player 1 *believes that player* 2 *thinks that the most probable contribution of player* 1 *is all his endowment. Using*

$$f_1^2(x) = \int_{\theta=0}^{\theta=2} f_1^2(x|\theta)\, \pi_1^2(\theta)\, d\theta, \tag{3.9}$$

then (3.7), along with (3.2) and (3.3), imply the unconditional density:

$$f_1^2(0) = 0,\ f_1^2(x) = (\ln 2)\, x - x \ln x,\ x \in (0, 2], \tag{3.10}$$

and, hence, the unconditional cumulative distribution:

$$F_1^2(0) = 0,\ F_1^2(x) = \frac{1}{4}x^2 + \frac{1}{2}(\ln 2)\, x^2 - \frac{1}{2}x^2 \ln x,\ x \in (0, 2]. \tag{3.11}$$

Of course, had we used (3.8) instead of (3.7), in conjunction with (3.2), (3.3) and (3.9), we would have got unconditional distributions different from (3.10) and (3.11).

3.4. Consistency of beliefs and actions

In a psychological Nash equilibrium, beliefs and actions are consistent with each other (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009). However, we do not require consistency of beliefs and actions. Furthermore, such a consistency is often violated empirically. Hence, the relevant distributions and the signal θ_i are taken to be given exogenously. In this respect, we take an empirically based modelling strategy that is identical to the one followed in Ellingsen et al. (2010), Dufwenberg et al. (2011) and Khalmetski et al. (2015).

3.5. Psychological utility functions

We shall specify and discuss three utility functions for three different groups of individuals depending on the information available to them; these three groups belong to APR1, APR2 and PUB. We shall compare each term in the utility function for one group with the analogous term in the other two groups.

3.5.1. Psychological utility for the APR treatment

The *psychological utility function* of a player 1 in group APR1 is given by (3.12), below, and the *psychological utility function* of a player 2 in group APR2 is given by (3.13), below.

$$\begin{aligned} U_1^{APR}(g_1, g_2, \theta_1) &= u_1(g_1, g_2) + \left[\phi_1^S(g_1, \theta_1) + \phi_1^I(g_1)\right] + \kappa_1 R_1, & (3.12)\\ U_2^{APR}(g_2, g_1) &= u_2(g_2, g_1) + \left[\phi_2^S(g_2) + \phi_2^I(g_2)\right] + \kappa_2 R_2, & (3.13) \end{aligned}$$

where $\kappa_1, \kappa_2 \geq 0$. Player 1 (who is in group APR1) is the informed player, and he receives a signal, θ_1, about what player 2 expects him to contribute. Player 2 (who is in group APR2), the uninformed partner, receives no signal. Hence, the utility of player 1, in (3.12), depends on θ_1 but the utility of player 2, in (3.13), does not depend on any signal.

Each of the utility functions, (3.12) and (3.13), has three terms on the RHS. The first terms $u_1(g_1, g_2)$ and $u_2(g_2, g_1)$, respectively, are the same (material) utility functions as in the classical public goods game, (2.1). The second terms in each utility function capture the emotions of guilt and surprise–seeking. The third terms in each utility function reflect concerns for reciprocity. We now explain the second and third terms in detail and give precise specifications for them.

3.5.2. Guilt-aversion and surprise-seeking

Let

$$\nu_i \in [0,1], \alpha_i \geq 0, \beta_i \geq 0, i = 1, 2. \tag{3.14}$$

Assuming that player 1 has a degree of empathy for player 2, it is reasonable to assume that player 1 gains utility from positively surprising player 2, but suffers a utility loss by negatively surprising player 2. This is formalized by the function $\phi_1^S(g_1, \theta_1)$ in (3.12) above, and (3.15) below.

$$\phi_1^S(g_1, \theta_1) = \nu_1 \left\{ \alpha_1 \left[\int_{x=0}^{g_1} (g_1 - x) f_1^2(x|\theta_1)\, dx \right] - \beta_1 \left[\int_{x=g_1}^{y} (x - g_1) f_1^2(x|\theta_1)\, dx \right] \right\}. \tag{3.15}$$

Ex-ante, player 2 expects player 1 to contribute $x \in [0, y]$ with probability density $f_2^1(x)$. But player 1 does not know $f_2^1(x)$. Instead, player 1 forms a second order belief, with probability density $f_1^2(x)$, about player 2's expectation of the contribution, g_1, of player 1. Player 1 is the informed player and he receives a signal, θ_1, from player 2. Thus, he uses the conditional density $f_1^2(x|\theta_1)$. Ex-post, player 2 discovers that player 1 has actually contributed $g_1 \in [0, y]$. For $x \in [0, g_1]$, player 1 expects player 2 to be pleasantly surprised. This contributes positive utility to player 1. Thus, player 1 is *surprise seeking*. For $x \in [g_1, y]$, player 1 expects player 2 to be disappointed. This contributes negative utility to player 1, possibly because he suffers guilt for disappointing player 2, i.e., player 1 is *guilt-averse*.[12] Thus, $\phi_1^S(g_1, \theta_1)$ is called *the simple surprise function for player* 1.[13]

Analogously, $\phi_2^S(g_2)$, in (3.13) above, and (3.16) below, is *the simple surprise function for player* 2. Note that $\phi_2^S(g_2)$ does not depend on a signal. This is because, since player 2 is the uninformed player, he does not receive any signal to condition on.

[12] A player suffers disutility if he thinks he has negatively surprised his partner. Yet, maybe surprisingly, he himself does not suffer disutility from a negative surprise inflicted on him by his partner. A term that captures the latter is introduced in subsection 10.4, Appendix D. And similarly for positive surprises and the intentions behind positive and negative surprises. These extra terms, however, do not change any of our results. Therefore, we have omitted them to simplify the exposition.

[13] This function was first introduced by Khalmetski et al. (2015).

$$\phi_2^S(g_2) = \nu_2 \left\{ \alpha_2 \left[\int_{x=0}^{g_2} (g_2 - x) f_2^2(x)\, dx \right] - \beta_2 \left[\int_{x=g_2}^{y} (x - g_2) f_2^2(x)\, dx \right] \right\}. \quad (3.16)$$

Assuming that player 1 has a degree of empathy for player 2, it is reasonable to assume that player 1 gains utility from believing that player 2 thinks that player 1 intended to positively surprise him but suffers a utility loss from believing that player 2 thinks that player 1 intended to negatively surprise him.[14] This is formalized by the function $\phi_1^I(g_1)$ in (3.12) above, and (3.17) below.[15]

$$\phi_1^I(g_1) = (1 - \nu_1) \left\{ \alpha_1 \left[\int_{x=0}^{g_1} (g_1 - x) f_1^4(x)\, dx \right] - \beta_1 \left[\int_{x=g_1}^{y} (x - g_1) f_1^4(x)\, dx \right] \right\}. \quad (3.17)$$

Player 2 believes, with probability density $f_2^3(x)$, that player 1 thinks that player 2 expects player 1 to contribute $x \in [0, y]$. But player 1 does not know $f_2^3(x)$. Instead, player 1 forms a fourth order belief, with probability density $f_1^4(x)$, about player 2's belief that player 1 thinks that player 2 expects player 1 to contribute $x \in [0, y]$. Comparing (3.15) and (3.17), the only difference is in the distributions used. Thus, the two terms in the brackets in (3.17) are, respectively, the surprise–seeking and guilt–aversion tendencies, when taking into account the role of intentions. For this reason, $\phi_1^I(g_1)$ is called *the attribution of intentions function for player* 1.

Ex-post, once the experiment is complete, player 2 can observe the contribution, g_1, of player 1, and update his third order beliefs, f_2^3, which in turn gives rise to conditional fourth order beliefs of player 1, f_1^4 (i.e., $f_2^3(x \mid g_1)$ and $f_1^4(x \mid g_1)$). However, in this paper, we suppress the dependence on g_1. With appropriate conditions, for instance, to ensure that the second order condition holds, the qualitative results in our paper go through. The reasonableness and the empirical soundness of these extra conditions could perhaps be explored more fully in future research.[16]

Analogously, $\phi_2^I(g_2)$, in (3.13) and (3.18), is *the attribution of intentions function for player* 2.

$$\phi_2^I(g_2) = (1 - \nu_2) \left\{ \alpha_2 \left[\int_{x=0}^{g_2} (g_2 - x) f_2^4(x)\, dx \right] - \beta_2 \left[\int_{x=g_2}^{y} (x - g_2) f_2^4(x)\, dx \right] \right\}. \quad (3.18)$$

Finally, note that for the surprise function for player 1, $\phi_1^S(g_1, \theta_1)$ in (3.12) and (3.15), above, we conditioned on θ_1, the signal player 1 received about what player 2 expected him to contribute. However, for the attribution of intentions function for player 1, $\phi_1^I(g_1)$

[14] Suppose I stepped on your toe. This is, of course, painful to you and, therefore, psychologically painful to me. Furthermore, suppose that I believed that you thought that my action was deliberate rather than accidental. Then, my belief would increase my psychological pain.

[15] Following Khalmetski et al. (2015), α_1 and β_1 in (3.15) and (3.17) are identical.

[16] We are very grateful to a referee for pointing this out.

in (3.12) and (3.17), above, we *did not* condition on θ_1. This is because player 1 knows that player 2 does not know that player 1 has received the signal θ_1. Hence, the third order belief of player 2, f_2^3, does not depend on θ_1. In turn, the forth order belief of player 1, f_1^4, which is a belief about f_2^3, is also independent of θ_1.

3.5.3. Reciprocity

We now consider the final terms on the RHS of (3.12) and (3.13).

Consider player 1. Following Dufwenberg and Kirchsteiger (2004), we define the reciprocity term in (3.12) as[17]

$$R_1 = k_{12}\widehat{k}_{21}, \tag{3.19}$$

where k_{12} is the *kindness of the first player to the second, as perceived by player 1* and $\widehat{k}_{21}$ is the *kindness of player 2 to player 1, as perceived by player 1.* This is the sense in which reciprocity is conditional. If player 2 is perceived to be kind ($\widehat{k}_{21} > 0$), then by reciprocating the kindness ($k_{12} > 0$), player 1 increases utility, given in (3.12). Similarly, utility can be increased by reciprocating unkindness ($\widehat{k}_{21} < 0$) with unkindness ($k_{12} < 0$).

The computation of k_{21} (kindness of player 2 to player 1) requires the specification of an *equitable utility* to player 1, u_1^E. Player 2 does not observe g_1 so he must use his first order beliefs, b_2^1, in determining u_1^E.

$$u_1^E = \int_{x=0}^{y} \left[\frac{1}{2}\max\left\{u_1\left(x, g_2\right),\, g_2 \in [0, y]\right\} + \frac{1}{2}\min\left\{u_1\left(x, g_2\right),\, g_2 \in [0, y]\right\}\right] f_2^1\left(x\right) dx, \tag{3.20}$$

where $u_1\left(x, g_2\right)$ is defined in (2.1). The equitable utility, u_1^E, is an equally weighted average of the maximum and the minimum utilities that player 2 can guarantee player 1 through the contribution decision, g_2. Given the definition of $u_1\left(g_1, g_2\right)$ in (2.1), the highest possible material utility to player 1 arises when $g_2 = y$ and the lowest when $g_2 = 0$. Thus, we can rewrite (3.20) as

$$u_1^E = \textstyle\int_{x=0}^{y} v_1\left(y - x\right) f_2^1\left(x\right) dx + r\bar{b}_2^1 + r\dfrac{y}{2}, \tag{3.21}$$

where $\bar{b}_2^1$, the average first order belief of player 2 about the contribution of player 1, is given by

$$\bar{b}_2^1 = \textstyle\int_{x=0}^{y} x f_2^1\left(x\right) dx. \tag{3.22}$$

We now define k_{21} as follows.

$$k_{21} = Eu_1 - u_1^E, \tag{3.23}$$

[17] The kindness functions in Rabin (1993) and Dufwenberg and Kirchsteiger (2004) are related in spirit, although the specifications are slightly different.

where, the expected utility of player 1, Eu_1, is given by

$$Eu_1 = \int_{x=0}^{y} u_1(x, g_2)\, f_2^1(x)\, dx, \tag{3.24}$$

and $u_1(x, g_2)$ is defined in (2.1). From (3.23), player 2 is kind to player 1 if through the choice of a contribution, g_2, player 1 receives expected material utility greater than the equitable utility. Otherwise player 2 is unkind to player 1. Substituting (3.21), (3.24) in (3.23), we get

$$k_{21}(g_2) = rg_2 - \frac{1}{2}ry. \tag{3.25}$$

If player 1 cares about reciprocity, then he needs to form inferences about $k_{21}(g_2)$ when making his contribution decision, g_1. Since player 1 does not observe g_2, he uses his first order beliefs about g_2, given by b_1^1. Thus, his perception of the kindness of player 2 is $\widehat{k}_{21}(b_1^1) = \int_{x=0}^{y} k_{21}(x)\, f_1^1(x)\, dx$, or

$$\widehat{k}_{21}\left(b_1^1\right) = r\bar{b}_1^1 - \frac{1}{2}ry, \tag{3.26}$$

where $\bar{b}_1^1$ denotes the average first order belief of player 1 about the contributions of player 2, i.e.,

$$\bar{b}_1^1 = \int_{x=0}^{y} x f_1^1(x)\, dx. \tag{3.27}$$

Next, we compute the kindness of player 1 to player 2 as perceived by player 1, k_{12}. We first need to compute the equitable utility of player 2, u_2^E. Proceeding as in the computation of (3.25), we have

$$k_{12}(g_1) = rg_1 - \frac{1}{2}ry. \tag{3.28}$$

Substituting (3.26) and (3.28) in (3.19) we get

$$R_1 = R_1(g_1, b_1^1) = r^2\left(g_1 - \frac{y}{2}\right)\left(\bar{b}_1^1 - \frac{y}{2}\right). \tag{3.29}$$

By an analogous argument, R_2, in (3.13), is given by

$$R_2 = R_2(g_2, b_2^1) = r^2\left(g_2 - \frac{y}{2}\right)\left(\bar{b}_2^1 - \frac{y}{2}\right). \tag{3.30}$$

From (3.29) and (3.30), on account of reciprocity, the players would like to contribute more than half the endowment if they expect the other player to contribute, on average, more than half the endowment.

3.5.4. Psychological utility for the PUB treatment

Recall that in PUB, each player, i, receives a signal, θ_i, about the contribution, g_i, that his partner, player $-i$, expects him (player i) to make. Furthermore, each player i knows that

his partner, player $-i$, has received that signal and this is public knowledge. It follows that the densities that enter the psychological utility function for player i in PUB are conditional on θ_i. Hence, the psychological utility function of player i in PUB is given by:

$$U_i^{PUB}(g_i, g_{-i}, \theta_i) = u_i(g_i, g_{-i}) + \phi_i^S(g_i, \theta_i) + \phi_i^I(g_i, \theta_i) + \kappa_i R_i, \tag{3.31}$$

where the functions $\phi_i^S(g_i, \theta_i)$ and $\phi_i^I(g_i, \theta_i)$ are given by:

$$\phi_i^S(g_i, \theta_i) = \nu_i \left\{ \alpha_i \left[\int_{x=0}^{g_i} (g_i - x) f_i^2(x|\theta_i)\, dx \right] - \beta_i \left[\int_{x=g_i}^{y} (x - g_i) f_i^2(x|\theta_i)\, dx \right] \right\}, \tag{3.32}$$

$$\phi_i^I(g_i, \theta_i) = (1 - \nu_i) \left\{ \alpha_i \left[\int_{x=0}^{g_i} (g_i - x) f_i^4(x|\theta_i)\, dx \right] - \beta_i \left[\int_{x=g_i}^{y} (x - g_i) f_i^4(x|\theta_i)\, dx \right] \right\}, \tag{3.33}$$

and the parameters are as in (3.14) above. The reciprocity term R_i, $i = 1, 2$, is identical to (3.29), (3.30).

The interpretation of (3.31), (3.32) and (3.33) is the same as for (3.12) to (3.18) except for the introduction of the conditioning on θ_i.

Of particular interest is the difference between $\phi_1^I(g_1)$ and $\phi_2^I(g_2)$ on the one hand and $\phi_i^I(g_i, \theta_i)$ on the other hand. As explained above in detail, $\phi_1^I(g_1)$ and $\phi_2^I(g_2)$ depend on the unconditional fourth order beliefs of the two players, $f_1^4(x)$ and $f_2^4(x)$, respectively, in the APR treatment. In contrast $\phi_i^I(g_i, \theta_i)$, in the PUB treatment, depends on the conditional fourth order beliefs $f_i^4(x|\theta_i)$.

3.6. Psychological equilibria

Recall from Section 3.4 that, as in Khalmetski et al. (2015) and Ellingsen et al. (2010), we do not force consistency of beliefs and actions. Also recall the description of the APR and the PUB treatments (Subsection 3.2). This allows us to state the definitions of psychological equilibria in the two treatments.

Definition 1 *: A psychological equilibrium for the APR treatment is a pair of contributions, $(\widehat{g}_1, \widehat{g}_2) \in [0, y]^2$, such that $\widehat{g}_1$ maximizes player 1's psychological utility (3.12) given $\widehat{g}_2$, the distributions f_1^2, f_1^4 and the signal $\theta_1 \in [0, y]$; and $\widehat{g}_2$ maximizes player 2's psychological utility (3.13) given $\widehat{g}_1$ and the distributions f_2^2, f_2^4.*

Definition 2 *: A psychological equilibrium for the PUB treatment is a pair of contributions, $(g_1^*, g_2^*) \in [0, y]^2$, such that g_1^* maximizes player 1's psychological utility (3.31), with $i = 1$ and $-i = 2$, given g_2^*, the distributions f_1^2, f_1^4 and the signal $\theta_1 \in [0, y]$; and g_2^* maximizes player 2's psychological utility (3.31), with $i = 2$ and $-i = 1$, given g_1^*, the distributions f_1^2, f_1^4 and the signal $\theta_2 \in [0, y]$.*

Notation: Recall that we have denoted by $\widehat{g}_i$ and g_i^*, the optimal contributions under, respectively, the APR and the PUB treatments. We shall use $\widetilde{g}_i$ to refer to either $\widehat{g}_i$ or g_i^* when no distinction need be made.

Definition 3 *(Dominant actions): In the psychological equilibrium, $(\widetilde{g}_1, \widetilde{g}_2)$, $\widetilde{g}_1$ is a dominant action for player 1 if $\widetilde{g}_1$ maximizes player 1's psychological utility for any given $g_2 \in [0, y]$ (not just $\widetilde{g}_2$). Likewise, $\widetilde{g}_2$ is a dominant action for player 2 if $\widetilde{g}_2$ maximizes player 2's psychological utility for any given $g_1 \in [0, y]$ (not just $\widetilde{g}_1$).*

4. Theoretical Predictions

In this section we derive the theoretical predictions of our model (all proofs are in the Appendix). Our assumptions on the continuity of the objective function and the compactness of the constraint set ensures that an equilibrium exists. Furthermore, the next proposition shows that the equilibrium is in dominant actions.[18]

Proposition 2 *: A psychological equilibrium exists, and is in dominant actions.*

A simple condition on the relative importance of the two psychological tendencies of surprise–seeking and guilt–aversion ensures that the equilibrium is unique. This condition is strongly borne out by our empirical results.

Proposition 3 *: If guilt–aversion is not less important than surprise–seeking ($\alpha_i \leq \beta_i$), then $\widetilde{g}_i$ is unique.*

In the next proposition, we consider the comparative static results with respect to the preference parameters α_1, β_1 and α_2, β_2 which denote the relative importance of surprise–seeking and guilt–aversion in the utility functions of players. Both tendencies push in the direction of greater contributions (see (3.15), (3.16)). An increase in α_i increases the propensity to surprise the partner by exceeding the partner's expectations; this induces higher contributions. An increase in β_i increases guilt from falling below the expectations of the partner; this too increases contributions.

Proposition 4 *(Comparative statics with respect to α_i and β_i) Consider an interior solution at which the second order condition strictly holds. Then, at this interior solution, the following results hold.*
(a) Informed players in the APR treatment:
(i) $\frac{\partial \widehat{g}_1}{\partial \alpha_1} \geq 0$ and $\frac{\partial \widehat{g}_1}{\partial \beta_1} \geq 0$,

[18] The result on equilibrium in dominant actions follows from the quasi-linear structure of preferences, which are typically employed in the public goods game literature.

(ii) $\frac{\partial \widehat{g}_1}{\partial \alpha_1} > 0$ *for* $\nu_1 > 0$ *and* $F_1^2(\widehat{g}_1|\theta_1) > 0$, *or* $\nu_1 < 1$ *and* $F_1^4(\widehat{g}_1) > 0$.
(iii) $\frac{\partial \widehat{g}_1}{\partial \beta_1} > 0$ *for* $\nu_1 > 0$ *and* $F_1^2(\widehat{g}_1|\theta_1) < 1$, *or* $\nu_1 < 1$ *and* $F_1^4(\widehat{g}_1) < 1$.
(b) Uninformed players in the APR treatment:
(i) $\frac{\partial \widehat{g}_2}{\partial \alpha_2} \geq 0$ *and* $\frac{\partial \widehat{g}_2}{\partial \beta_2} \geq 0$,
(ii) $\frac{\partial \widehat{g}_2}{\partial \alpha_2} > 0$ *for* $\nu_2 > 0$ *and* $F_2^2(\widehat{g}_2) > 0$, *or* $\nu_2 < 1$ *and* $F_2^4(\widehat{g}_2) > 0$.
(iii) $\frac{\partial \widehat{g}_1}{\partial \beta_2} > 0$ *for* $\nu_2 > 0$ *and* $F_2^2(\widehat{g}_2) < 1$, *or* $\nu_2 < 1$ *and* $F_2^4(\widehat{g}_2) < 1$.
(c) Players in the PUB treatment:
(i) $\frac{\partial g_i^*}{\partial \alpha_i} \geq 0$ *and* $\frac{\partial g_i^*}{\partial \beta_i} \geq 0$,
(ii) $\frac{\partial g_i^*}{\partial \alpha_i} > 0$ *for* $\nu_i > 0$ *and* $F_i^2(g_i^*|\theta_i) > 0$, *or* $\nu_i < 1$ *and* $F_i^4(g_i^*|\theta_i) > 0$.
(iii) $\frac{\partial g_i^*}{\partial \beta_i} > 0$ *for* $\nu_i > 0$ *and* $F_i^2(g_i^*|\theta_i) < 1$, *or* $\nu_i < 1$ *and* $F_i^4(g_i^*|\theta_i) < 1$.

How does player i alter contributions based on the received signal, θ_i (this rules out uninformed players in the APR treatment)? It turns out that the answer to this question is critical in separating the relative importance of surprise–seeking and guilt–aversion.

Proposition 5 *: (Comparative statics with respect to* θ_i *under Assumption A4) Consider an interior solution at which the second order condition strictly holds. Then, at this interior solution, the following results hold.*
(a) Informed players in the APR treatment: For $\nu_1 = 0$, $\frac{\partial \widehat{g}_1}{\partial \theta_1} = 0$, *and for* $\nu_1 > 0$, $\frac{\partial \widehat{g}_1}{\partial \theta_1} \gtreqless 0 \Leftrightarrow \alpha_1 \lesseqgtr \beta_1$,
(b) Players in the PUB treatment: $\frac{\partial g_i^*}{\partial \theta_i} \gtreqless 0 \Leftrightarrow \alpha_i \lesseqgtr \beta_i$, $i = 1, 2$.

Proposition 5 states that contributions are an increasing (decreasing) function of the signal if, and only if, guilt aversion is relatively more (less) important than surprise seeking. Testing this proposition requires observing the contribution decision of players for different signals, which can be achieved with the strategy method. This leads to the construction of our within–subjects design, as in Khalmetski et al. (2015), that we describe in Section 5 below.

Proposition 6 *(Reciprocity): Consider an interior solution at which the second order condition strictly holds. Assume that* $\kappa_i > 0$. *Optimum contributions in the APR treatment,* $\widehat{g}_i$, *and in the PUB treatment,* g_i^*, *are increasing in the average first order beliefs of the player,* $\overline{b}_i^1$, *about the contributions of the other player,* $i = 1, 2$.

Proposition 6 brings out the role of reciprocity; if a player believes that the other player will contribute a greater amount, on average, then the player reciprocates by contributing more.

Proposition 7 *: Suppose that* $\alpha_1 \leq \beta_1$. *If intentions are unimportant* ($\nu_1 = 1$), *then* $\widehat{g}_1 = g_1^*$.

According to Proposition 7, if intentions are unimportant ($\nu_1 = 1$), then the contribution of an informed player in the asymmetric private treatment (APR1) is identical to the contribution of that same player in the public treatment (PUB). We shall see in Subsection 6.3 that this (equality of contributions in the treatments APR1 and PUB) is rejected by the evidence; hence, simple surprise–seeking and simple guilt–aversion are insufficient to explain the evidence. In particular, the attribution of intentions functions $\phi_1^I(g_1)$ and $\phi_1^I(g_1, \theta_1)$ are also important (recall Subsection 3.5).

Remark 1 *: Proposition 7 gives a sufficient, but not necessary, condition for $\widehat{g}_1 = g_1^*$. Thus, if $\widehat{g}_1 \neq g_1^*$ for a particular player, then we can infer that intentions are important for that player.*[19] *The converse (if $\widehat{g}_1 = g_1^*$, then intentions are unimportant) does not hold because, from (3.33), ϕ_i^I has two terms and the marginal effect arising from one may cancel the marginal effect of the other. This remark will play an important role in interpreting our experimental findings. Under more general preferences, this result might not hold.*

If intentions are unimportant then (and only then), the choice relevant terms in the utility function under the PUB treatment and the APR1 treatment are identical; letting $\nu_1 = 1$, compare (3.12) with (3.31) (ϕ_1^I is given, respectively, in (3.17) and (3.33)). Assuming that guilt–aversion is more important than surprise–seeking, then the optimum (in both cases) is unique (Proposition 3). Hence, the contribution has to be the same for both, APR1 and PUB. However, this is no more the case with APR2. The choice relevant terms in APR2 are not the same as under PUB, even when intentions are unimportant. This is because players in PUB can observe a signal of the other player's expectation when forming their simple surprise function, while players in APR2 do not observe any signal; setting $\nu_2 = 1$, compare (3.13), (3.18) with (3.31), (3.33) for $i = 2$. So, we cannot say anything, in general, about the level of contributions under APR2 and PUB, even for a player under identical information conditions in stage 1 of our experiments (see subsections 5.1 and 5.2). It all depends on the specifics of the probability distributions.

When $g_1^* \neq \widehat{g}_1$, our model allows for both cases: $g_1^* < \widehat{g}_1$ and $g_1^* > \widehat{g}_1$; this is shown in the next proposition.

Proposition 8 *: Suppose $\alpha_1 < \beta_1$, $\nu_1 < 1$, and $g_1^* \in (0, y)$ or $\widehat{g}_1 \in (0, y)$. Then, there is a $\theta_1^* \in (0, y)$ such that $\theta_1 < \theta_1^* \Rightarrow g_1^* < \widehat{g}_1$ and $\theta_1 > \theta_1^* \Rightarrow g_1^* > \widehat{g}_1$.*

[19] This conclusion requires that all players do indeed maximize the psychological utility functions of the form that we have proposed. If this is not the case, then the contributions in the APR1 and the PUB treatment can also differ because of other motivations that we have not considered in our paper.

5. Within-subjects experimental design

We consider two treatments in our within–subjects design: The *asymmetric private treatment* (APR) and the *public treatment* (PUB).[20]

We use the method of induced beliefs as originally used in Ellingsen et al. (2010) and replicated in Khalmetski et al. (2015). Ellingsen et al. (2010) use a between–subjects design, while we use the within–subjects design of Khalmetski et al. (2015), which is the appropriate method to test Proposition 5. As in Ellingsen et al. (2010) and Khalmetski et al. (2015), we use the partner's guesses as the experimental measure of second order beliefs (SOB); this is a central feature, and main advantage, of the induced beliefs design. In order to compare our results with several earlier papers, we also employ the between–subjects design that is described in Section 7, below; however, it cannot be employed to test Proposition 5.

There were 222 subjects who participated in the within–subjects design and were randomly matched in pairs to play the public goods game. Subjects were undergraduate students in Qingdao Agriculture University in China and they belonged to a cross section of disciplines. The initial endowment of each player was 20 tokens (1.5 tokens equal 1 Yuan).

To control for possible order effects, we ran the two treatments in a counterbalanced order. In our *Experiment 1*, all subjects participated in the APR treatment first, followed by the PUB treatment. This order was reversed in *Experiment 2*. A total of 108 subjects participated in Experiment 1 and 114 subjects participated in Experiment 2. No subjects participated in both experiments. Across both treatments, we obtained 7104 data points.

In order to minimize the possibility of biasing the responses of subjects, they played the APR and the PUB treatments before learning about the outcomes from the treatment that they played first. After having played both treatments, one of the two treatments was chosen randomly and played for real money with the subjects; this ensures incentive-compatibility of the experimental design.

5.1. Asymmetric private treatment (APR)

The APR treatment, which is described in detail in Appendix-B, has the following stages.

Stage-1: Subjects are initially asked to guess their partner's possible contribution to

[20] We did not have the *symmetric private treatment* in our within-subjects design but we have such a treatment in our between-subjects design that is described in Section 7. The downside of the symmetric treatment is that some subjects may infer that their guesses could also be obtained by their partners. Deception is not allowed in economic experiments, so we could not have lied to the subjects that their partner is not informed about their guess. The asymmetric treatment is not subject to this potential criticism.

the public good on a *Guess Sheet* that allows guesses from zero to 20 tokens.[21] The guesses were incentivized in all our treatments and designs (see, for instance, Appendix B).

Stage-2: After the Guess Sheets are collected, the subjects receive the *Decision Sheet* that implements the strategy method in our within–subjects design. The information-advantageous group, APR1, received the following instruction: "Your partner doesn't know that you will be informed about his/her guess, and s/he is not informed about your guess". This enables us to exclude the possibility that some subjects in group APR1 may suspect that their guesses may be revealed to their partners.

Using the *strategy method*, the decision sheets for the APR1 subjects (player 1) required them to decide on their actual contribution, $g_1 \in [0, 20]$, for each possible value of the signal, $\theta_1 \in \{0, 1, 2, ..., 20\}$, received from the partner (player 2). This gives 21 data points for each member of APR1.

APR2 subjects, unlike APR1 subjects, are not informed that their guesses could not be obtained by their partners. Nor do we use the strategy method with APR2 subjects. Rather, an APR2 subject (player 2) makes a contribution, g_2, based on his/her beliefs; for instance, if APR2 subjects suffer from guilt-aversion, then they use their second order beliefs, b_2^2, of the partner's first order belief, b_1^1, about g_2.

Stage-3: If the APR treatment (from among APR and PUB) is chosen at the end of the experiment to be played for real money, then each informationally advantageous subject (player 1) is informed of the partner's guess (θ_1 from the Guess Sheet of Stage-1). Using the partner's actual guess, each player's contribution (g_1, g_2) is determined accordingly to the contribution decision already made in the Decision Sheet in Stage-2. Once each player's contributions are determined in this manner, the outcome is implemented.

5.2. Public treatment

In the public treatment, PUB, the first stage is identical to the APR treatment described in Subsection 5.1. In the second stage, however, players have to decide on a level of contribution for each possible public announcement of the first order belief of the other player. The subjects received instructions for Stage 2 only after Stage 1 was completed (as in APR1 treatment). Each player is told: "Your partner is also informed about your guess of his/her contribution before s/he decides to contribute. And s/he is informed that you know his/her guess before you choose your contribution." The provision of this information distinguishes the PUB treatment from the APR treatment in the following respect. Each

[21] Instead of asking subjects to guess the *average* contribution of all other subjects, we asked each subject to guess the contribution of their *single* partner. The reason is that the expectation of the average contribution might serve as the norm to some extent, and this might consequently raise subjects' aversion from deviating from falling below the norm. However, the aim of our experiments is to investigate the existence of guilt-aversion that arises from contributing below the other player's expectation.

player i, $i = 1, 2$, can condition on the signals, respectively, θ_i and θ_{-i} (and not just one of the players and one of the signals), and both players know this.

6. Results and discussion of the within–subjects design

6.1. Testing Proposition 5: Surprise-seeking or guilt–aversion?

Proposition 5 allows us to distinguish between the relative strengths of surprise–seeking and guilt–aversion. We regress the contributions of a player (as revealed by the strategy method) on the guesses of the other player for each information-advantageous individual.[22] Recall that the information-advantageous subjects decide on their contributions, conditional on knowing the guesses of the partners; these guesses correspond to the signal θ_1 received by player 1 in our model. Hence the guesses/signals may be taken as an important summary statistic of their second order beliefs. This is the distinguishing feature of the induced beliefs method, which is also employed by Ellingsen et al. (2010) and Khalmetski et al. (2015).

The resulting distribution of the regression coefficients that are significant at the 5% level is shown in Figure 6.1.[23] In Figure 6.1, 5% of the subjects exhibit negative coefficients, and the remaining 95% of the subjects exhibit positive coefficients; Proposition 5 predicts that the former are relatively surprise–seeking, while the latter are relatively guilt-averse. Therefore, most of our subjects are relatively guilt-averse.

The average size of the negative coefficients is -0.71, and the average size of the positive coefficients is 0.74. A t-test of differences in means is precluded because the negative case has less than 10 observations. The two distributions of positive and negative coefficients are not significantly different ($p = 0.000$ for a two-sided Mann-Whitney U test). Excluding the two-direction changing cases where contributions are non-monotonic in guesses, the average sizes of the negative and positive coefficients is, respectively, -1 and 0.89 ($p = 0.000$ for a two-sided Mann-Whitney U test comparing the two distributions of coefficients).

The within–subjects regressions in the dictator game experiments of Khalmetski et al. (2015) show that surprise–seeking plays a relatively larger role as compared to our results. In their analogue of Figure 6.1, about 70% of the coefficients are distributed to the right of

[22] Introducing further regressors, e.g., gender, education, field of study creates perfect multicollinearity. The reason for this is that our strategy method contains 21 contribution decisions for each subject in group APR1. For each subject his/her demographic characteristics are always the same. See the decision sheet in Appendix B.

[23] The proportion of subjects that have significant regression coefficients is $81/111 = 73\%$. If we consider all subjects, including those who have insignificant regression coefficients, then the proportion of negative coefficients (surprise seeking) is 9%, while the proportion of positive coefficients (guilt aversion) is 91%. This result is similar to the case where we compare only the significant coefficients.

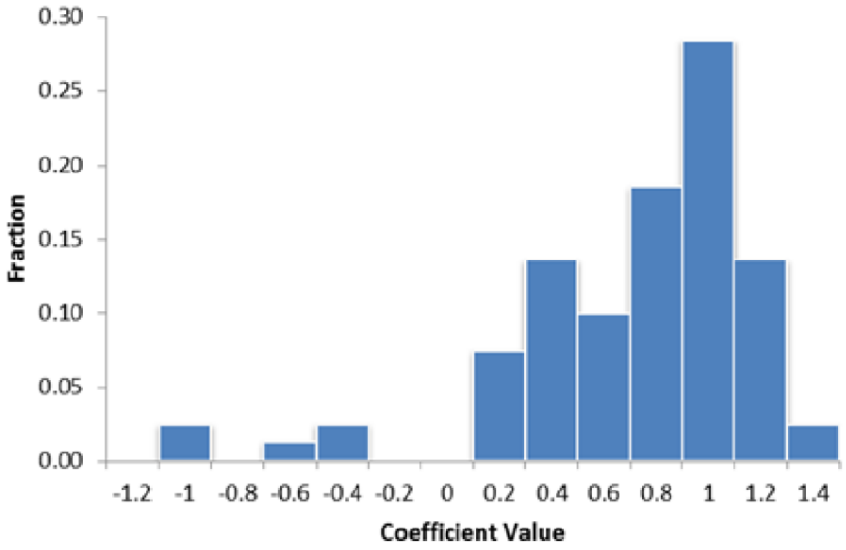

Figure 6.1: The distribution of regression coefficients of contributions on guesses (second order beliefs) that are significant at the 5% level in a within-subjects regression.

zero (compare this to 95% positive coefficients in Figure 6.1). In conjunction, these results indicate that guilt–aversion is more important than surprise–seeking for most subjects; though more so for the public goods game than for the dictator game.

In our experiments, across all subjects, contributions and second order beliefs have a strong positive and significant correlation. In order to ensure statistical independence across observations, for each individual we use only the actually realized SOB/contribution pair.[24] The Spearman correlation coefficient is 0.41 and 0.36, respectively, in the APR1 and PUB treatments; $p = 0.001$ in both treatments. This is an important finding of our paper. It shows that in a strategic setting with the induced beliefs method and a within–subjects design, the results on the importance of guilt–aversion are consistent (at the individual and at the aggregate level) with earlier results that used neither within–subjects nor the induced beliefs method (Dufwenberg and Gneezy, 2000; Dufwenberg et al., 2011; Guerra and Zizzo, 2004; Reuben et al., 2009).

Indeed the finding of zero overall correlation between actions and second order beliefs that has been found using induced beliefs in a between–subjects design (Ellingsen et al., 2010), and a within–subjects design (Khalmetski et al., 2015), using dictator games, does not generalize to the public goods game. Fehr and Schmidt (2006) assert, based on the evidence, that perhaps the results from the dictator game have a special status that is not always transferable to other strategic contexts.

Ellingsen et al. (2010) also report a zero correlation between actions and second order

[24] We are grateful to a referee's suggestion to proceed in this manner.

	Experiment 1			Experiment 2		
	APR1	APR2	PUB	APR1	APR2	PUB
FOB	15.39	15.39	14.31	12.81	14.65	13.36
	(77.0)	(77.0)	(71.6)	(64.1)	(73.3)	(66.8)
Contribution	12.05	13.20	10.74	11.90	13.06	12.19
	(60.3)	(66.0)	(53.7)	(59.5)	(65.3)	(61.0)

Table 6.1: Average first order belief (FOB) and the average contributions. Note: Figures in parentheses give the percentage of contributions relative to the endowment of 20 tokens. In the APR treatment, the information-advantageous group is labelled as APR1, while the rest are labelled by APR2.

beliefs for a trust game. However, they use a between–subjects design and not, unlike us and Khalmetski et al. (2015), a within–subjects design and the strategy method. In light of these findings, perhaps the original challenge that was perceived for models of guilt–aversion, and psychological game theory in general, based on the findings of Ellingsen et al. (2010), now appears to have a narrower scope.

6.2. Testing Proposition 6: The role of reciprocity

From Proposition 6, we expect that reciprocity will induce a positive correlation between average first order beliefs, $\bar{b}_i^1$, of player i and his/her level of contributions, g_i. We do not observe $\bar{b}_i^1$, but we believe that it is eminently plausible that the observed first order beliefs of the player, b_i^1, are likely to be positively correlated with the average beliefs, $\bar{b}_i^1$. Hence, in this paper, we take b_i^1 as a proxy for $\bar{b}_i^1$, for the purposes of a statistical analysis between first order beliefs and contributions.

The Spearman correlation coefficient between FOB and contributions is 0.30 ($p = 0.001$) in APR1, and 0.37 ($p = 0.000$) in PUB, which is significantly positive at the 1% level in both cases. Therefore, the contribution choices of our subjects were at least partly driven by reciprocity, i.e., if player 1 believed that player 2 is likely to contribute a large amount, then player 1 reciprocates by contributing more. These results are supported further by our regression analysis below.

6.3. Testing Proposition 7: The importance of intentions

Proposition 7 states that if intentions are unimportant ($\nu_1 = 1$), then $\widehat{g}_1 = g_1^*$. Thus, if $\widehat{g}_1 \neq g_1^*$, then intentions are important (Recall Remark 1). We now test this prediction.

Table 6.1 shows the summary statistics of the first stage guesses of the players (the first order beliefs, denoted by FOB) and the contributions of players in both treatments (APR and PUB) in Experiments 1 and 2.[25] From Table 6.1, we see that contributions range from

[25]The terms 'Experiment 1' and 'Experiment 2' are defined in Section 5.

	Relatively Higher in PUB	Relatively Higher in APR1	Row Total
Experiment 1	13.0%	20.3%	33.3%
Experiment 2	15.8%	12.3%	28.1%
Column Average	14.4%	16.2%	30.6%

Table 6.2: The proportions reported in the second column of the table comprise the subsets for which the mean contributions under PUB are relatively higher at 5 percent. In the third column, the proportions comprise the subsets for which the mean contributions under APR1 are relatively higher at 5 percent. The row total and column average are also shown.

59.5% to 65.3% of the endowment. These figures are much higher than in the dictator game experiments that use the induced beliefs method. For instance, in Khalmetski et al. (2015), dictators gave 23% of their endowments to recipients; the corresponding figure for Ellingsen et al. (2010) is 24%. Also from Table 6.1, we see that FOBs range from 64.1% to 77%. In Khalmetski et al. (2015), the average first order belief was 34% of the endowment; the corresponding figure for Ellingsen et al. (2010) is 32%.

Recall that for each subject, we have 21 conditional contribution decisions for each of the APR1 and PUB treatments. A two-sided Mann-Whitney U test between the two distributions, contributions in APR1 and contributions in PUB, at the individual-level, revealed the following results at the 5% significance level. In Experiment 1, 17 out of 54 subjects made significantly different contribution decisions in the two treatments; the corresponding figure for Experiment 2 is 16 out of 57 subjects. To determine further, the treatment in which contributions are higher, we used the t-test to compare the means of the two distributions. We use these results to report, in Table 6.2, the proportion of individuals for whom contributions in one treatment were significantly higher at 5%.

In our experiments, 30.6% of subjects across both experiments exhibited $\widehat{g}_1 \neq g_1^*$. Thus, intentions were important for, at least, 30% of our subjects. Of these, 14.4% exhibited $\widehat{g}_1 < g_1^*$ and 16.2% exhibited $\widehat{g}_1 > g_1^*$. As noted in Proposition 8, both behaviors are consistent with our model.

In models of purely inequity averse individuals, the beliefs of other players should not influence the contributions of players. By contrast, in the APR treatment, around 91.9% of the information-advantageous subjects changed their conditional contributions at least once (in the PUB treatment, that we consider in Subsection 6.3, this figure is 90.5%). For 73.9% of the subjects, the within–subjects correlation of the contributions of players with the guesses of their partner is significant at the 5% level. These results support a central assumption of psychological game theory, namely, that beliefs directly influence actions (in a manner that goes beyond simple Bayesian updating).

The response of contributions to changes in the beliefs of the partner is also sharper in public goods experiments relative to dictator game experiments. For instance, in the

	APR1	APR2	PUB
FOB	0.011	0.268	0.156
Contribution	0.708	0.750	0.230

Table 6.3: p-values in Mann-Whitney U Tests.

only other directly comparable study, that of Khalmetski et al. (2015): (1) 77.5% of the dictators changed their transfers at least once in response to a change in the guesses of the other player (the corresponding figure in our study is 91.9%). (2) For 53.9% of the dictators, the within–subjects correlation of transfers with guesses is significant at the 5% level (the corresponding figure in our study is 73.9%).

6.4. Are order effects important?

Consider the order effects which distinguish Experiments 1 and 2.

Table 6.3 shows the p-values in a two-sided Mann-Whitney U test.[26] The null hypothesis is that the distribution of FOB/contributions is not different in the two experiments. Only the p-value of the FOB of APR1 is less than the 5% significance level. Hence, other than the distribution of the information-advantageous group's FOB, there are no significant order effects in contributions or in the FOB.

7. Empirical tests using a between–subjects design

In this section, we describe the findings from our between–subjects design, while continuing to use the induced beliefs method. This allows us to compare our results with the closely related study of Khalmetski et al. (2015), and with our findings from the within–subjects design. Furthermore, the induced beliefs findings of Ellingsen et al. (2010) arose in a between–subjects design, although they did not use the PUB treatment. In this section, we also compare their results with ours.

We use the following three treatments in the between–subjects design: the *private treatment* (PR), the *asymmetric private treatment* (APR), and the *public treatment* (PUB). The treatments APR and PUB are similar to those described in the within–subjects design, except that in a between–subjects design we do not use the strategy method (recall this was needed to test Proposition 5). Thus, we elicit a level of contribution from each player for a single guess of the other player, rather than their underlying strategy for each possible guess of the other player. For instance, in the PUB treatment in the between–subjects design, before a player makes the contribution decision, the screen display contains the

[26] We used only one observation per subject (the actually realized belief/contribution pair) to conduct the MWU tests, hence, the observations are independent.

Contribution	0	1	2	3	4	5	6	7	8	9	10	Total
PR	17	8	10	6	7	15	10	5	5	2	17	102
	(16.7)	(7.8)	(9.8)	(5.9)	(6.9)	(14.7)	(9.8)	(4.9)	(4.9)	(2.0)	(16.6)	
APR1	7	1	2	1	9	7	2	3	7	0	14	53
	(13.2)	(1.9)	(3.8)	(1.9)	(17.0)	(13.2)	(3.8)	(5.7)	(13.2)	(0.0)	(26.3)	
APR2	10	0	2	2	7	6	3	2	5	1	15	53
	(18.9)	(0.0)	(3.8)	(3.8)	(13.2)	(11.3)	(5.7)	(3.8)	(9.4)	(1.9)	(28.2)	
PUB	13	6	7	7	4	23	8	8	5	0	19	100
	(13.0)	(6.0)	(7.0)	(7.0)	(4.0)	(23.0)	(8.0)	(8.0)	(5.0)	(0.0)	(19.0)	
Total	47	15	21	16	27	51	23	18	22	3	65	308

Table 7.1: The frequencies of contributions in the between-subjects design. Note: Figures in parentheses give the percentage of the subjects making the associated contributions.

information of their partner's guess of their contribution. The following information is displayed on the computer screen: "Your partner is also informed about your guess of his/her contribution before s/he decides to contribute. And s/he is informed that you know his/her guess before you choose your contribution".

The treatment PR is identical to the APR treatment except that there is no information advantageous group that is given special instructions (see Stage-2 of the APR treatment in Subsection 5.1).[27] In our between–subjects design, the existing pool of players is randomly paired; each pair plays the game only once. The experimental design closely follows Ellingsen et al. (2010) and Khalmetski et al. (2015).

Our subjects are undergraduate and postgraduate students in Nankai University and Tianjin University (China). There are 18 sessions that are split equally between the three treatments (6 sessions per treatment).[28] A total of 308 subjects took part in the experiment, and nobody attended more than one session. The initial endowment was set at 10 tokens (1 token = 1.5 Yuan).

The frequency distribution of contributions, $0, 1, ..., 10$, for the full between–subjects dataset is shown in Table 7.1. The results for each of the three treatments are described separately below. In each case, we replicate the result in Ellingsen et al. (2010) with induced beliefs and direct elicitation of contributions. Namely, the correlation between contributions and second order beliefs is not significantly different from zero at the 5%

[27] Recall that we do not have a PR treatment in the within-subjects design.

[28] Three sessions of the private treatment and three sessions of the public treatment were run in December 2014. The remaining sessions were run in March-April 2015. To examine if there was a temporal effect arising from the two different dates of the sessions, we compared the contribution and beliefs in the two different set of sessions. The Mann-Whitney U tests show that there is no significant difference in (1) the private treatment (for the contributions comparison, $p = 0.489$ and for the beliefs comparison, $p = 0.811$), and (2) the public treatment (for the contributions comparison, $p = 0.672$ and for the beliefs comparison, $p = 0.668$). All the APR sessions were run on one date only, so there were no issues of timing.

level except for the APR treatment where it is significant at the 10% level.[29] However, the regression analysis in Section 7.4 shows that second order beliefs are a significant determinant of contributions, hence, guilt–aversion is important after all.

7.1. Private treatment (PR)

Of the 6 standard private sessions, four had 18 subjects each, one had 14 subjects, and one had 16 subjects[30]. In total, we obtained 102 observations. The average contribution is 4.63 tokens out of an endowment of 10 tokens, and the average second order belief is 5.03 tokens. On average, the subjects expect others to contribute about 0.40 tokens more than the actual contribution (two-sided t-test, $p = 0.357$). About 16.7% of the subjects contribute nothing, and about 16.6% contribute the entire endowment. 37% subjects contribute more than (or equal to) the signal that they receive of their partner's beliefs.

7.2. Asymmetric private treatment (APR)

In the 6 sessions for the APR treatment, each session had 18 subjects, except one session which had 16 subjects. In total, there are 106 observations. The average contribution is 5.73 tokens out of an endowment of 10 tokens. About 16% of the subjects contribute nothing, and about 27% contribute the entire endowment. 59% of the subjects contribute more than (or equal to) their own second order belief.

The average contribution of APR1 subjects (information-advantageous player) is 5.81 tokens, and the average belief is 6.08 tokens. Hence, on average, these subjects expect about 0.27 tokens more than the real contribution (two-sided t-test, $p = 0.652$). The average contribution of the non-information-advantageous subjects is 5.64 tokens.

The average contribution of the information-advantageous subjects is 1.18 tokens higher than that of the subjects in the PR treatment (two-sided t-test, $p = 0.044$). The contribution distributions of subjects in the PR treatment and the information-advantageous subjects in the APR treatment are significantly different from each other at the 10% level (two-sided Mann-Whitney U Test, $p = 0.057$). In contrast, a non-parametric test of the comparison of distributions of the contributions of subjects in the PR treatment and contributions of non information-advantageous subjects in the APR treatment shows that they are not significantly different (two-sided Mann-Whitney U Test, $p = 0.122$). Overall, significantly positive correlation is found between contributions and (induced) second order beliefs.

[29] The figures for the APR treatment are as follows: Pearson coefficient = 0.244, $p = 0.078$; Spearman coefficient = 0.239, $p = 0.085$.

[30] The variation in the sessions arose from no-shows, although each session had 18 subjects signed-in.

7.3. Public treatment (PUB)

In the 6 sessions in the PUB treatment, there were 18 subjects in four sessions and 14 subjects each in the remaining two sessions, giving a total of 100 observations. The average contribution was 5.06 tokens out of an endowment of 10 tokens, and the average second order belief was 5.91 tokens. Hence, the subjects expect about 0.85 tokens more than the actual contributions of other players (two-sided t-test, $p = 0.051$). Only 13 subjects contribute nothing in all the sessions, while about 19 contribute the entire endowment; 64 subjects contribute more than or equal to their own second order belief.

7.4. Determinants of contributions

We now consider the determinants of contributions that include beliefs and individual level characteristics of the subjects such as gender, experience in similar experiments, and field of study. We ran several Tobit models to explore such effects; see Table 7.2.[31]

The variables FOB and SOB denote, respectively, first and second order beliefs of a subject. The explanatory variable 'Education' takes values from the set $\{1, 2, ..., 7\}$ with higher values denoting higher educational attainment (e.g., Education = 1 for first year undergraduate students and Education = 6 for second year master students). The dummy variable 'Male' equals 1 for male, and 0 for female. The dummy variable 'Field of Study' equals 1 if the subject studies economics or business and zero otherwise. The dummy variable 'Experience' equals 1 if the subject has attended similar experiments before. The treatment variable is a dummy variable, D_{PUB}. In Models 1, 2 and 3, D_{PUB} equals 1 for the PUB treatment, and 0 for the PR and APR1 treatments; while in Models 4, 5 and 6, D_{PUB} equals 1 for the PUB treatment, and 0 for the APR1 treatment. We also considered a range of interaction terms.

The variables SOB and FOB have significant effects on the contribution decision in almost all models. FOB and SOB are positively and significantly correlated with the contribution; this reflects, respectively, reciprocity, and guilt from falling below the contributions of the other player. However, the interaction term $SOB \times D_{PUB}$ did not reveal any significant effect (so it is not reported in Table 7.2). The FOB is positively correlated with contributions; this captures feelings of reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004). Since our SOB are induced beliefs, they are different from the

[31] Our implementation of the Tobit models is similar to that in Khalmetski et al. (2015). The Tobit models account for the share of observations with zero contributions and those with contributions of 10 tokens. Khalmetski et al. (2015) account for zero contributions; our results are similar if we account for zero contributions alone or zero and 10 tokens. Additionally, our Tobit model can allow for clustered standard errors which deal with the potential heteroskedasticity across different experimental sessions and the intra-session correlation. The OLS results are similar in terms of the magnitudes and the significance of the coefficients, so we have not reported them here but these are available from the authors on request.

Dependent Variable	Contribution					
Tobit Model	1	2	3	4	5	6
D_{PUB}	0.07	−0.05	3.22	−1.13	−1.07	2.96
	[0.645]	[0.639]	[3.814]	[0.886]	[0.870]	[3.768]
SOB		0.22**	0.40**		0.26*	0.39**
		[0.105]	[0.174]		[0.132]	[0.170]
FOB			1.12***			1.09***
			[0.235]			[0.228]
Education			−0.23			−0.25
			[0.603]			[0.591]
Male			−0.01			−0.06
			[1.255]			[1.236]
Field (of study)			−2.08			−1.88
			[1.434]			[1.475]
Experience			0.11			−0.05
			[2.695]			[2.690]
Male×D_{PUB}			−3.23**			−3.13**
			[1.522]			[1.506]
Other interactions			insig.			insig.
Constant	5.16***	3.96***	−1.70	6.35***	4.78***	−1.36
	[0.419]	[0.664]	[3.320]	[0.742]	[1.078]	[3.287]
Observations	255	255	255	153	153	153
Log-Likelihood	−599.8	−597.3	−527.4	−353.6	−351.4	−305.3

Table 7.2: Determinants of public good contributions. The dependent variable is individual-level contributions to the public good. All Tobit models are censored from both sides. Superscripts stars, ***, **, * denote significance levels of 1 percent, 5 percent, and 10 percent, respectively. Clustered standard errors in brackets (clustering on experimental sessions).

FOB. When SOB are self-reported (and not induced) there is likely to be a significant correlation between SOB and FOB (Dufwenberg et al. 2011). This is not an issue in our study, hence, both kinds of beliefs retain statistical significance in our model. This, and the lowest value for the log-likelihood for model 6, suggest that reciprocity and guilt aversion, jointly, explain best the contribution decisions of players. Male subjects tended to contribute significantly less than female subjects in the PUB treatment; gender has been shown to be an important determinant of economic decisions elsewhere (Croson and Gneezy, 2009; Eckel and Grossman, 2008; Gneezy and Rustichini, 2004).

There were no significant difference between the contributions of economics/business students and others, which separates these results from some others on social preferences (Fehr et al., 2006). Previous experience of participating in similar experiments does not significantly affect the contribution decisions.

The differences in aggregate contributions in the PUB treatment relative to the APR and PR treatments, as captured by the dummy variable D_{PUB}, is not statistically significant. This result stands in contrast to the dictator game results of Khalmetski et al. (2015) who found that aggregate dictator giving in their public treatment was significantly higher relative to the their private treatment.

8. Reconciling the extent of guilt-aversion/surprise-seeking in different experiments

As discussed above (see, for instance, Table 6.1), in our within-subjects design, we find that most subjects, 95%, are guilt-averse and only 5% are surprise-seeking. This contrasts with the corresponding 70%–30% mix in Khalmetski et al. (2015). How might these findings be reconciled? There are two potential competing responses.

1. Differences in the nature of the games: Only one player has a non-trivial action in the dictator game, while in the public goods game both players have non-trivial actions. The public goods game may also invoke heuristics of mutual cooperation that are absent in the dictator game. On the other hand, the dictator in the dictator game may feel a sense of entitlement to his own endowment. Thus, surprise-seeking may be more important in dictator games than in public goods game. Conversely, guilt-aversion may be more important in public good games than in dictator games. Ultimately, these remarks are also linked to the issue of portability of dictator game experiments that is increasingly well documented (Fehr and Schmidt, 2006; Dhami, 2016). Indeed, these differences might be sufficient to explain the stated problem, but this is a conjecture that we cannot test with our data.

2. Reciprocity: The reviewers of this paper, however, offer another plausible explana-

tion.[32] Essentially, the idea is that reciprocity, which can be exercised in public goods games, but not in dictator games, may be a potential explanation. However, to rigorously formulate this idea, additional assumptions have to be made, none of which can be tested by the data in our experiments. The remaining part of this section is focussed on developing the reciprocity explanation.

So far, we have assumed that the first order belief, b_i^1, does not depend on the signal θ_i. Specifically, $F_i^1(x|\theta_i) = F_i^1(x)$, $i = 1, 2$ (Assumption A4). With the exception of Proposition 5, none of our results depend on how b_i^1 depends on θ_i. We now relax Assumption A4. The new assumption, Assumption A5, requires that when player i observes a higher signal, θ_i, from player $j \neq i$, then player i assigns a lower probability that player j will make a low contribution. Formally:

Assumption A5 : (a) For a player in APR1, $F_1^1(x|\theta_1)$ is differentiable in θ_1 and $\frac{\partial F_1^1(x|\theta_1)}{\partial \theta_1} < 0$ for all $x \in (0, y)$ and all $\theta_1 \in (0, y)$.
(b) For a player in PUB, $F_i^1(x|\theta_i)$ is differentiable in θ_i and $\frac{\partial F_i^1(x|\theta_i)}{\partial \theta_i} < 0$ for all $x \in (0, y)$ and all $\theta_i \in (0, y)$, $i = 1, 2$.

We may give the following intuition for Assumption A5.[33] A higher first-order belief of player 2 (signaled by θ_1) is assumed to imply that player 2 is more likely to make a higher contribution. The justification is that if player 2 expects that player 1 is going to make a high contribution, then player 2's reciprocity may push him to be kind in response (as also suggested by Proposition 6). If player 1 anticipates this motivation of player 2, then after player 1 observes a higher signal θ_1 about the FOB of player 2, he would anticipate that player 2 is going to make a higher contribution (i.e., the FOB of player 1 would be higher). This basically corresponds to the statement of Assumption A5.

The only terms affected by the change from Assumption A4 to Assumption A5 are those involving the distribution of first order beliefs, F_1^1 in the APR1 treatment and F_i^1, $i = 1, 2$, in the PUB treatment (players in the APR2 treatment receive no signal, θ_1, and do not realize that players in the APR1 treatment receive a signal, θ_1). All other terms are unaffected.[34] Specifically, the unconditional distributions, $f_1^1(x)$ in the APR1 treatment and $f_i^1(x)$, $i = 1, 2$, in the PUB treatment are replaced by the conditional distributions $f_1^1(x|\theta_1)$ and $(f_i^1|\theta_i)$, $i = 1, 2$. Thus, the analogue of (3.27), in the case of APR1 and PUB, is given by:

[32] We are grateful in particular to Reviewer 2, who has offered a particularly well formulated statement of his/her views. Below, we attempt to rigorously formulate the reviewers' suggestion.

[33] We are grateful to a referee for suggesting the writing of this text.

[34] Referee 2 points out that players may be subject to the false consensus effect in ex-ante reporting their first order beliefs about the other player. While we do control for the false consensus effect with respect to second order beliefs in our experiments, we do not control for such an effect in our first order beliefs.

$$\bar{b}_i^1(\theta_i) = \int_{x=0}^{y} x f_i^1(x|\theta_i)\, dx;\ i = 1 \text{ for APR1 and } i = 1,2 \text{ for PUB.}$$

The reciprocity of player i (the analogue of (3.29) and (3.30)) is given by

$$R_i(g_i, b_i^1, \theta_i) = r^2 \left(g_i - \frac{y}{2}\right)\left(\bar{b}_i^1(\theta_i) - \frac{y}{2}\right);\ i = 1 \text{ for APR1 and } i = 1,2 \text{ for PUB} \quad (8.1)$$

Using integration by parts, we can write $\bar{b}_i^1(\theta_i) = y - \int_{x=0}^{y} F_i^1(x|\theta_i)\, dx$. Thus, the derivative of $\bar{b}_i^1(\theta_i)$ can be written as

$$\frac{d\bar{b}_i^1(\theta_i)}{d\theta_i} = -\int_{x=0}^{y} \frac{\partial F_i^1(x|\theta_i)}{\partial \theta_i} dx > 0, \quad (8.2)$$

where the sign follows from Assumption A5. This captures the central implication of Assumption A5. A higher signal, θ_i, increases the average belief that player i has about the contribution of the partner. This insight modifies the result in Proposition 5; we state the modified result next.

Proposition 9 *(Comparative statics with respect to θ_i under Assumption A5) Suppose Assumption 5 holds. Consider an interior solution at which the second order condition strictly holds. Then, at this interior solution, the following results hold.*
(a) Informed players in the APR treatment (APR1):
(i) For $\nu_1 = 0$ and $\kappa_1 = 0$: $\frac{\partial \widehat{g}_1}{\partial \theta_1} = 0$.
(ii) For $\nu_1 = 0$ and $\kappa_1 > 0$: $\frac{\partial \widehat{g}_1}{\partial \theta_1} > 0$.
(iii) For $\nu_1 > 0$:

$$\frac{\partial \widehat{g}_1}{\partial \theta_1} \gtreqless 0 \Leftrightarrow \alpha_1 \gtreqless \beta_1 + \frac{d\bar{b}_1^1(\theta_1)}{d\theta_1} \frac{\kappa_1 r^2}{\nu_1 \left|\frac{\partial F_1^2(g_1|\theta_1)}{\partial \theta_1}\right|}. \quad (8.3)$$

(b) Players in the PUB treatment:

$$\frac{\partial g_i^*}{\partial \theta_i} \gtreqless 0 \Leftrightarrow \alpha_i \gtreqless \beta_i + \frac{d\bar{b}_i^1(\theta_1)}{d\theta_i} \frac{\kappa_i r^2}{\nu_i \left|\frac{\partial F_i^2(g_i|\theta_i)}{\partial \theta_i}\right| + (1-\nu_i)\left|\frac{\partial F_i^4(g_i|\theta_i)}{\partial \theta_i}\right|},\ i = 1,2. \quad (8.4)$$

Using Assumption A3 and (8.2), the RHS of the last inequality in each of (8.3) and (8.4) is positive. In the case of APR1 it is strictly positive if $\kappa_1 > 0$, while for PUB it is strictly positive if $\kappa_i > 0$.

We can now reconcile our finding that most subjects, 95%, in our public good game are guilt-averse and only 5% are surprise-seeking, in contrast with the corresponding 70%–30% mix in the Khalmetski et al. (2015) dictator game.

Recall that if Assumption A4 is correct then, from Proposition 5, we have for the PUB treatment $\frac{\partial g_i^*}{\partial \theta_i} \gtreqless 0 \Leftrightarrow \alpha_i \lesseqgtr \beta_i$ and that for the APR1 treatment (if $\nu_1 > 0$) $\frac{\partial \widehat{g}_1}{\partial \theta_1} \gtreqless 0 \Leftrightarrow \alpha_1 \lesseqgtr \beta_1$. Thus if a player exhibits $\frac{\partial \widehat{g}_1}{\partial \theta_1} > 0$ or $\frac{\partial g_i^*}{\partial \theta_i} > 0$, then we can conclude that that player is relatively more guilt-averse than surprise-seeking ($\alpha_i > \beta_i$).

On the other hand, if Assumption A5 (rather than A4) is correct, then, from Proposition 9, we could have a player who is relatively more surprise-seeking than guilt-averse ($\alpha_i > \beta_i$) still exhibit $\frac{\partial \widehat{g}_1}{\partial \theta_1} > 0$ (APR1 treatment, $\nu_1 > 0$ and $\kappa_1 > 0$) or $\frac{\partial g_i^*}{\partial \theta_i} > 0$ (PUB treatment and $\kappa_i > 0$).

Thus, our finding that players are more likely to be guilt-averse in our public goods game than in the dictator game could simply be caused by us mistakenly (on the basis of A4) counting some players who are relatively more surprise-seeking ($\alpha_i > \beta_i$) as guilt-averse.

9. Conclusions

Our aim in this paper is to make theoretical and empirical contributions to the literature on psychological game theory. We emphasize three different but possibly related emotions: (1) Reciprocity, (2) simple guilt–aversion/surprise–seeking, and (3) the attribution of intentions behind guilt–aversion/surprise–seeking.

The work by Ellingsen et al. (2010), using induced beliefs, called into question the very existence of guilt–aversion as a relevant emotion. We extend the theoretical framework of Khalmetski et al. (2015), which was developed for dictator games, to the public goods game which allows a role for strategic interaction.

Using an induced beliefs methodology, as in Ellingsen et al. (2010), we implement a within–subjects design with the strategy method, and a between–subjects design that does not employ the strategy method. Earlier research had used one or the other of these two designs, which sometimes creates difficulty in comparing the results.

In the within–subjects design, we find that, in the statistically significant cases, the vast majority of our subjects (95%) are relatively guilt-averse and only 5% are relatively surprise seeking; we offer a novel explanation based on a reciprocity channel (Section 8), for the differences in our results from those of previous studies that find relatively greater surprise-seeking. We also find guilt–aversion at the aggregate level. In contrast, Khalmetski et al. (2015) find no aggregate guilt aversion because guilt–aversion and surprise–seeking in individual data counteracted each other at the aggregate level.

In our between–subjects design, if we use only correlation analysis, we replicate the results of Ellingsen et al. (2010) of zero correlation between second order beliefs and actions. However, a regression analysis shows that second order beliefs have a significant effect on actions. Hence, guilt–aversion plays a statistically significant role in determining

contributions. However, the between–subjects design cannot distinguish between guilt–aversion and surprise–seeking. We find that for, at least, 30% of our subjects, attribution of intentions behind guilt–aversion/surprise–seeking is important. However, we cannot rule out this motive for our remaining subjects. Finally, reciprocity also helps explain public goods contributions, jointly with the other psychological factors that we consider.

In this paper we have used a relatively simple and tractable theoretical framework, which involved additive separability between the various psychological motivations and a standard quasilinear form for material utility. Future research might wish to relax these features of our model and explore, in richer detail, the resulting strategic issues. The choice between these alternative specifications is ultimately an empirical matter.

References

[1] al-Nowaihi, A. and Dhami, S., 2015. Evidential equilibria: Heuristics and biases in static games of complete information. Games 6, 637-676.

[2] Battigalli, P., and Dufwenberg, M., 2007). Guilt in games. American Economic Review 97, 170-176.

[3] Battigalli, P., and Dufwenberg, M. (2009). Dynamic psychological games. Journal of Economic Theory. 144(1), 1-35.

[4] Charness, G., and Dufwenberg, M. (2006). Promises and partnership. Econometrica. 74(6), 1579-1601.

[5] Croson, R., and Gneezy, U., 2009. Gender Differences in Preferences. Journal of Economic Literature 47, 1-27.

[6] Dhami, S., 2016. The foundations of behavioral economic analysis. Oxford: Oxford University Press.

[7] Dufwenberg, M., Gächter, S., and Hennig-Schmidt, H., 2011. The framing of games and the psychology of play. Games and Economic Behavior 73, 459-478.

[8] Dufwenberg, M., and Gneezy, U., 2000. Measuring Beliefs in an Experimental Lost Wallet Game. Games and Economic Behavior 30, 163-182.

[9] Dufwenberg, M., and Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games and Economic Behavior 47, 268-298.

[10] Eckel, C. C. and Grossman, P. J., 2008. Differences in the economic decisions of men and women, experimental evidence. In Plott, C. and Smith, V. (Eds.). Handbook of Experimental Economics Results, Volume 1. New York: Elsevier.

[11] Ellingsen, T., Johannesson, M. Tjøtta, S., Torsvik, G., 2010. Testing Guilt Aversion. Games and Economic Behavior 68, 95-107.

[12] Elster, J., 1989. Social Norms and Economic Theory. Journal of Economic Perspectives 3, 99-117.

[13] Elster, J., 1998. Emotions in Economic Theory. Journal of Economic Literature 36, 47-74.

[14] Fehr, E., Naef, M., and Schmidt, K.M., 2006. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. American Economic Review 96, 1912-1917.

[15] Fehr, E., and Schmidt, K., 2006. The economics of fairness, reciprocity and altruism, Experimental evidence and new theories. In Kolm S-C.and Ythier, J. M., (Eds.) Handbook of the Economics of Giving, Altruism and Reciprocity, Volume 1., New York: Elsevier.

[16] Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. Games and Economic Behavior 1, 60-79.

[17] Gneezy, U., and Rustichini, A., 2004. Gender and Competition at a Young Age. American Economic Review 94, 377-381.

[18] Guerra, G., and Zizzo, D. J., 2004. Trust Responsiveness and Beliefs. Journal of Economic Behavior and Organization 55, 25-30.

[19] Khalmetski, K., 2016. Testing guilt aversion with an exogenous shift in beliefs. Games and Economic Behavior 97, 110-119.

[20] Khalmetski, K., Ockenfels, A., Werner, P., 2015. Surprising Gifts. Journal of Economic Theory 159, 163-208.

[21] Rabin, M., 1993. Incorporating fairness into game theory and economics. American Economic Review 83, 1281-1302.

[22] Reuben, E., Sapienza, P., and Zingales, L., 2009. Is mistrust self-fulfilling? Economics Letters 104, 89-91.

[23] 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, 279-301.

[24] Vanberg, C., 2008. Why do people keep their promises? An experimental test of two explanations. Econometrica 76, 1476-1480.

10. Appendices

10.1. Appendix A : Proofs

Proof of Proposition 1: From (2.1), (2.2), $\frac{\partial u(g_i,g_{-i})}{\partial g_i} = r - v'(y-g_i) < 0$. Hence $(g_1^n, g_2^n) = (0,0)$. ■

Lemma 1 *: From (3.12)-(3.18), it follows that the utility of a player who is a member of APR1 can be written as*
$U_1^{APR}(g_1,g_2,\theta_1) = \Phi_1^{APR}(g_1,\theta_1) + rg_2$, *where*
$\Phi_1^{APR}(g_1,\theta_1) = v_1(y-g_1) + rg_1$
$+\nu_1\left\{\alpha_1\left[\int_{x=0}^{g_1}(g_1-x)\,f_1^2(x|\theta_1)\,dx\right] - \beta_1\left[\int_{x=g_1}^{y}(x-g_1)\,f_1^2(x|\theta_1)\,dx\right]\right\}$
$+(1-\nu_1)\left\{\alpha_1\left[\int_{x=0}^{g_1}(g_1-x)\,f_1^4(x)\,dx\right] - \beta_1\left[\int_{x=g_1}^{y}(x-g_1)\,f_1^4(x)\,dx\right]\right\} + \kappa_1 R(g_1,b_1^1)$,
is a function of g_1,θ_1 but not of g_2. Analogous expressions hold for members of APR2 and PUB.

Proof of Proposition 2: Consider a member of APR1. Given $g_2,\theta_1 \in [0,y]$, it follows from Lemma 1 that $\widehat{g}_1 \in [0,y]$ maximizes $U_1^{APR}(g_1,g_2,\theta_1)$ if, and only if, $\widehat{g}_1$ maximizes $\Phi_1^{APR}(g_1,\theta_1)$. Hence, such a $\widehat{g}_1$ will also maximize $U_1^{APR}(g_1,g_2,\theta_1)$ for *any* $g_2 \in [0,y]$. So, $\widehat{g}_1$ is a *dominant* action for player 1, if it exists. But it does exist because $[0,y]$ is compact and $\Phi_1^{APR}(g_1,\theta_1)$ is continuous in g_1. Similarly, player 1's partner from APR2 has a dominant action, $\widehat{g}_2$. Hence, $(\widehat{g}_1,\widehat{g}_2)$ is a psychological equilibrium, and is in dominant actions. Similarly, for the PUB treatment: A psychological equilibrium, (g_1^*,g_2^*), exists and is in dominant actions. ■

Lemma 2 *: Integrate the expression $\int_{x=0}^{g}(g-x)\,f(x)\,dx$ by parts, then differentiate, to get*
$\frac{\partial}{\partial g}\int_{x=0}^{g}(g-x)\,f(x)\,dx = F(g)$,
and, similarly,
$\frac{\partial}{\partial g}\int_{x=g}^{y}(x-g)\,f(x)\,dx = F(g)-1$.

Lemma 3 *: Consider a member of APR1. From Lemmas 1 and 2, it follows that:*
$\frac{\partial}{\partial g_1}U_1^{APR}(g_1,g_2,\theta_1) = r + \beta_1 - v_1'(y-g_1) + (\alpha_1-\beta_1)\left[\nu_1 F_1^2(g_1|\theta_1) + (1-\nu_1)F_1^4(g_1)\right] + \kappa_1 r^2\left(\bar{b}_1^1 - \frac{1}{2}y\right)$.
$\frac{\partial^2}{\partial g_1^2}U_1^{APR}(g_1,g_2,\theta_1) = v_1''(y-g_1) + (\alpha_1-\beta_1)\left[\nu_1 f_1^2(g_1|\theta_1) + (1-\nu_1)f_1^4(g_1)\right]$.
$\frac{\partial^2}{\partial g_1\partial\alpha_1}U_1^{APR}(g_1,g_2,\theta_1) = \nu_1 F_1^2(g_1|\theta_1) + (1-\nu_1)F_1^4(g_1)$.
$\frac{\partial^2}{\partial g_1\partial\beta_1}U_1^{APR}(g_1,g_2,\theta_1) = \nu_1\left[1-F_1^2(g_1|\theta_1)\right] + (1-\nu_1)\left[1-F_1^4(g_1)\right]$.
$\frac{\partial^2}{\partial g_1\partial\theta_1}U_1^{APR}(g_1,g_2,\theta_1) = \nu_1(\alpha_1-\beta_1)\frac{\partial F_1^2(g_1|\theta_1)}{\partial\theta_1}$.

Analogous expressions hold for APR2 (except that we do not condition on θ) and PUB. In particular, we note the following calculation for the PUB treatment for subsequent use.

$$\frac{\partial^2 U_i^{PUB}}{\partial g_i \partial \theta_i} = (\alpha_i - \beta_i)\left[\nu_i \frac{\partial F_i^2(g_i|\theta_i)}{\partial \theta_i} + (1-\nu_i)\frac{\partial F_i^4(g_i|\theta_i)}{\partial \theta_i}\right] - \kappa_i r^2 \int_{x=0}^{y} \frac{\partial F_i^1(x|\theta_i)}{\partial \theta_i} dx. \blacksquare$$

Proof of Proposition 3: Since $v_1'' < 0$, $\nu_1 \in [0,1]$, $f_1^2(g_1|\theta_1) \geq 0$, $f_1^4(g_1) \geq 0$, it follows, from Lemma 3, that $\frac{\partial^2}{\partial g_1^2} U_1^{APR}(g_1, g_2, \theta_1) < 0$ for $\alpha_1 \leq \beta_1$ and, hence, $\widehat{g}_1$ is unique. Analogous arguments show that $\widehat{g}_2$ and g_i^* are also unique. $\blacksquare$

Proof of Proposition 4: Similar to the proof of Proposition 5, below. $\blacksquare$

Proof of Proposition 5: Consider a member of APR1. By assumption, $0 < \widehat{g}_1 < y$ and $\frac{\partial^2}{\partial g_1^2} U_1^{APR}(\widehat{g}_1, g_2, \theta_1) < 0$. From the first of these, we get $\frac{\partial}{\partial g_1} U_1^{APR}(\widehat{g}_1, g_2, \theta_1) = 0$ and, hence, $\frac{\partial^2}{\partial g_1^2} U_1^{APR}(\widehat{g}_1, g_2, \theta_1) \frac{\partial \widehat{g}_1}{\partial \theta_1} = -\frac{\partial^2}{\partial g_1 \partial \theta_1} U_1^{APR}(g_1, g_2, \theta_1)$. From the second inequality, we get $sign \frac{\partial \widehat{g}_1}{\partial \theta_1} = sign \frac{\partial^2}{\partial g_1 \partial \theta_1} U_1^{APR}(g_1, g_2, \theta_1)$. Proposition 5(a) then follows from Lemma 3. Part (b) is similar. $\blacksquare$

Proof of Proposition 6: Using Lemma 3, and proceeding as in the proof of Proposition 5, we note that $sign \frac{\partial \widehat{g}_1}{\partial \overline{b}_1^1} = sign \frac{\partial g_1^*}{\partial \overline{b}_1^1} = \kappa_1 r^2 > 0$. $\blacksquare$

Proof of Proposition 7: Let $(\widehat{g}_1, \widehat{g}_2)$ be a psychological equilibrium of the APR treatment and let (g_1^*, g_2^*) be a psychological equilibrium of the PUB treatment. Suppose $\nu_1 = 1$ and assume that $\alpha_1 \leq \beta_1$. We want to show that $\widehat{g}_1 = g_1^*$. By Proposition 2, $\widehat{g}_1$ is a dominant action. Hence, $\widehat{g}_1$ also maximizes U_1^{APR} for $g_2 = g_2^*$ (not just $g_2 = \widehat{g}_2$). So, $\widehat{g}_1$ maximizes $U_1^{APR}(g_1, g_2^*, \theta_1)$ and g_1^* maximizes $U_1^{PUB}(g_1, g_2^*, \theta_1)$. However, for $\nu_1 = 1$, $U_1^{APR}(g_1, g_2^*, \theta_1) = U_1^{PUB}(g_1, g_2^*, \theta_1)$, from (3.12), (3.17), (3.31) and (3.33). From Proposition 3, we then get $\widehat{g}_1 = g_1^*$. $\blacksquare$

Proof of Proposition 8: Since $\alpha_1 < \beta_1$, guilt-aversion is more important than surprise-seeking, and $g_1^*, \widehat{g}_1$ exist and are unique (Propositions 2 and 3) and since $\nu_1 < 1$, fourth order beliefs are important (recall (3.12), (3.17), (3.31), (3.33)).

Consider the case $g_1^* \in (0, y)$. The case $\widehat{g}_1 \in (0, y)$ is similar. The objective function $U_1^{APR}(g_1, g_2, \theta_1)$ is given in Lemma 1. The objective function $U_1^{PUB}(g_1, g_2, \theta_1)$ can be written in a similar manner by changing the unconditional fourth order distribution $f_1^4(x)$ to a conditional distribution, $f_1^4(x|\theta_1)$. Differentiating $U_1^{APR}(g_1, g_2, \theta_1)$ and $U_1^{PUB}(g_1, g_2, \theta_1)$ with respect to g_1, we get

$$\frac{\partial U_1^{APR}}{\partial g_1} = r + \beta_1 - v_1'(y - g_1) + (\alpha_1 - \beta_1)\left[\nu_1 F_1^2(g_1|\theta_1) + (1-\nu_1) F_1^4(g_1)\right] + \kappa_1 r^2 \left(\overline{b}_1^1 - \frac{1}{2}y\right), \tag{10.1}$$

$$\frac{\partial U_1^{PUB}}{\partial g_1} = r + \beta_1 - v_1'(y - g_1) + (\alpha_1 - \beta_1)\left[\nu_1 F_1^2(g_1|\theta_1) + (1-\nu_1) F_1^4(g_1|\theta_1)\right] \quad (10.2)$$

$$+\kappa_1 r^2 \left(\bar{b}_1^1 - \frac{1}{2}y\right). \quad (10.3)$$

From (10.1) and (10.3), after simplification, we get

$$\frac{\partial U_1^{APR}}{\partial g_1} - \frac{\partial U_1^{PUB}}{\partial g_1} = (\beta_1 - \alpha_1)(1-\nu_1)\left[F_1^4(g_1|\theta_1) - F_1^4(g_1)\right]. \quad (10.4)$$

Since $g_1 = g_1^*$ maximizes $U_1^{PUB}(g_1, g_2, \theta_1)$, and since $g_1^* \in (0, y)$, we necessarily have $\left[\frac{\partial}{\partial g_1} U_1^{PUB}(g_1, g_2, \theta_1)\right]_{g_1 = g_1^*} = 0$. Hence, from (10.4), we get

$$\left[\frac{\partial}{\partial g_1} U_1^{APR}(g_1, g_2, \theta_1)\right]_{g_1 = g_1^*} = (\beta_1 - \alpha_1)(1-\nu_1)\left[F_1^4(g_1|\theta_1) - F_1^4(g_1)\right]. \quad (10.5)$$

By assumption, $\alpha_1 < \beta_1$ and $\nu_1 < 1$, hence, (10.5) gives

$$\left[\frac{\partial}{\partial g_1} U_1^{APR}(g_1, g_2, \theta_1)\right]_{g_1 = g_1^*} \gtreqless 0 \Leftrightarrow F_1^4(g_1^*|\theta_1) \gtreqless F_1^4(g_1^*). \quad (10.6)$$

Recall that $f_1^4(x)$ and $f_1^4(x|\theta_1)$, respectively, are the unconditional and conditional probability densities. Let $\pi_1(\theta_1)$ be the probability density of the prior belief of player 1 about θ_1, then

$$f_1^4(x) = \int_{\theta_1=0}^{\theta_1=y} f_1^4(x|\theta_1)\,\pi_1(\theta_1)\,d\theta_1. \quad (10.7)$$

Hence,

$$\begin{aligned} F_1^4(g_1^*) &= \int_{x=0}^{x=g_1^*} f_1^4(x)\,dx = \int_{x=0}^{x=g_1^*}\left[\int_{\theta_1=0}^{\theta_1=y} f_1^4(x|\theta_1)\,\pi_1(\theta_1)\,d\theta_1\right]dx \\ &= \int_{\theta_1=0}^{\theta_1=y}\left[\int_{x=0}^{x=g_1^*} f_1^4(x|\theta_1)\,dx\right]\pi_1(\theta_1)\,d\theta_1 = \int_{\theta_1=0}^{\theta_1=y} F_1^4(g_1^*|\theta_1)\,\pi_1(\theta_1)\,d\theta_1 \end{aligned}$$

We have that $\pi_1(\theta_1) \geq 0$, $\int_{\theta_1=0}^{\theta_1=y}\pi_1(\theta_1)\,d\theta_1 = 1$ and $F_1^4(g_1^*|\theta_1)$ is a continuous function of θ_1 (from Assumption A2), it follows, from the mean value theorem of definite integrals[35], that

$$\text{there is a } \theta_1^* \in (0, y), \text{ such that } F_1^4(g_1^*|\theta_1^*) = F_1^4(g_1^*). \quad (10.8)$$

From (3.1), which itself was a consequence of Assumptions A2 and A3, it follows that

$$\frac{\partial F_1^4(g_1^*|\theta_1)}{\partial \theta_1} < 0 \text{ for all } \theta_1 \in (0, y). \quad (10.9)$$

[35] We are using the following theorem. If $g : [a, b] \to R$ is continuous and h is an integrable function that does not change sign on $[a, b]$, then there exists $c \in (a, b)$ such that $\int_a^b g(x)\,h(x)dx = g(c)\int_a^b h(x)dx$.

From (10.8) and (10.9), we get that there is a $\theta_1^* \in (0, y)$, such that

$$\theta_1 < \theta_1^* \Rightarrow F_1^4(g_1^*|\theta_1) > F_1^4(g_1^*), \tag{10.10}$$

$$\theta_1 > \theta_1^* \Rightarrow F_1^4(g_1^*|\theta_1) < F_1^4(g_1^*). \tag{10.11}$$

From (10.6), (10.10) and (10.11) we get that there is a $\theta_1^* \in (0, y)$, such that

$$\theta_1 < \theta_1^* \Rightarrow \left[\frac{\partial}{\partial g_1} U_1^{APR}(g_1, g_2, \theta_1)\right]_{g_1 = g_1^*} > 0, \tag{10.12}$$

$$\theta_1 > \theta_1^* \Rightarrow \left[\frac{\partial}{\partial g_1} U_1^{APR}(g_1, g_2, \theta_1)\right]_{g_1 = g_1^*} < 0. \tag{10.13}$$

Suppose $\theta_1 < \theta_1^*$. From (10.12) we see that increasing g_1 beyond g_1^* increases utility, U_1^{APR}. But $g_1 = \widehat{g}_1$ maximizes utility U_1^{APR}. Hence, $\widehat{g}_1 > g_1^*$. Similarly, if $\theta_1 > \theta_1^*$ then $\widehat{g}_1 < g_1^*$. Hence, we have established that

$$\theta_1 < \theta_1^* \Rightarrow g_1^* < \widehat{g}_1,$$
$$\theta_1 > \theta_1^* \Rightarrow g_1^* > \widehat{g}_1. \blacksquare$$

Proof of Proposition 9 (Comparative statics with respect to θ_i under Assumption A5):

(a) Using the implicit function theorem, and Lemma 3, we get

$$\frac{\partial \widehat{g}_1}{\partial \theta_1} = \frac{(\alpha_1 - \beta_1)\,\nu_1 \frac{\partial F_1^2(g_1|\theta_1)}{\partial \theta_1} + \kappa_1 r^2 \frac{d\bar{b}_1^1(\theta_1)}{d\theta_1}}{-\frac{\partial^2 U_1^{APR}}{\partial g_1^2}}. \tag{10.14}$$

(i) For $\nu_1 = \kappa_1 = 0$, (10.14) gives $\frac{\partial \widehat{g}_1}{\partial \theta_1} = 0$.

(ii) For $\nu_1 = 0$, $\kappa_1 > 0$, (10.14) gives $\frac{\partial \widehat{g}_1}{\partial \theta_1} > 0$ since, by assumption, $\frac{\partial^2 U_1^{APR}}{\partial g_1^2} < 0$ and $\frac{\partial F_1^1(x|\theta_1)}{\partial \theta_1} < 0$.

(iii) Recall that, by assumption, $\frac{\partial^2 U_1^{APR}}{\partial g_1^2} < 0$, $\frac{\partial F_1^2(g_1|\theta_1)}{\partial \theta_1} < 0$ and $\frac{d\bar{b}_1^1(\theta_1)}{d\theta_1} > 0$. Hence, for $\nu_1 > 0$, (10.14) gives $\frac{\partial \widehat{g}_1}{\partial \theta_1} \gtreqless 0 \Leftrightarrow (\alpha_1 - \beta_1)\,\nu_1 \frac{\partial F_1^2(g_1|\theta_1)}{\partial \theta_1} + \kappa_1 r^2 \frac{d\bar{b}_1^1(\theta_1)}{d\theta_1} \gtreqless 0 \Leftrightarrow \alpha_1 - \beta_1 + \frac{\kappa_1 r^2 \frac{d\bar{b}_1^1(\theta_1)}{d\theta_1}}{\nu_1 \frac{\partial F_1^2(g_1|\theta_1)}{\partial \theta_1}} \gtreqless 0 \Leftrightarrow \alpha_1 \gtreqless \beta_1 + \frac{d\bar{b}_1^1(\theta_1)}{d\theta_1} \frac{\kappa_1 r^2}{\nu_1 \left|\frac{\partial F_1^2(g_1|\theta_1)}{\partial \theta_1}\right|}$.

(b) Using the implicit function theorem, and Lemma 3, we get

$$\frac{\partial g_i^*}{\partial \theta_i} = \frac{(\alpha_i - \beta_i)\left[\nu_i \frac{\partial F_i^2(g_i|\theta_i)}{\partial \theta_i} + (1 - \nu_i)\frac{\partial F_i^4(g_i|\theta_i)}{\partial \theta_i}\right] + \kappa_i r^2 \frac{d\bar{b}_i^1(\theta_1)}{d\theta_i}}{-\frac{\partial^2 U_i^{PUB}}{\partial g_i^2}}. \tag{10.15}$$

Since $\nu_i \in [0,1]$, $\frac{\partial F_i^2(g_i|\theta_i)}{\partial \theta_i} < 0$, $\frac{\partial F_i^4(g_i|\theta_i)}{\partial \theta_i} < 0$, $\frac{d\bar{b}_i^1(\theta_1)}{d\theta_i}$ and $\frac{\partial^2 U_i^{PUB}}{\partial g_i^2} < 0$, (10.15) gives:

$\frac{\partial g_i^*}{\partial \theta_i} \gtreqless 0 \Leftrightarrow (\alpha_i - \beta_i)\left[\nu_i \frac{\partial F_i^2(g_i|\theta_i)}{\partial \theta_i} + (1-\nu_i)\frac{\partial F_i^4(g_i|\theta_i)}{\partial \theta_i}\right] + \kappa_i r^2 \frac{d\bar{b}_i^1(\theta_1)}{d\theta_i} \gtreqless 0 \Leftrightarrow \alpha_i \lesseqgtr \beta_i + \frac{d\bar{b}_i^1(\theta_1)}{d\theta_i} \frac{\kappa_i r^2}{\nu_i\left|\frac{\partial F_i^2(g_i|\theta_i)}{\partial \theta_i}\right| + (1-\nu_i)\left|\frac{\partial F_i^4(g_i|\theta_i)}{\partial \theta_i}\right|}$. ■

10.2. Appendix B: Experimental instructions for the within-subject design (translation from Chinese instructions)

General information on the experiment

You are now participating in an economic experiment. If you read the following explanations carefully, you may be able to earn some money depending on your decisions and the decisions of others. During the experiment you are not allowed to communicate with other participants in any way. If you have questions, please raise your hand, and the experimenter will come to your desk.

During the experiment, we will not talk about Chinese Yuan, but about **tokens**. Your total income will first be calculated in tokens. The total amount of tokens that you have accumulated during the experiment will be converted into Chinese Yuan in cash at the end of the experiment at an exchange rate of **1.50 tokens = 1 Yuan**. Additionally, you will receive 5 Yuan, as a show-up fee for participating in this experiment. The experiment will be carried out only **once**.

The experiment consists of **two** parts.[36] First, you shall receive the instructions for the first part of the experiment. After the first part is completed, you shall receive the instructions for the second part of the experiment. After the experiment is completed, one part will be chosen randomly to be the payoff-relevant part. Each part consists of the *Guess Your Partner's Contribution Decision* and the *Contribution Decision*; this is explained below. In each part, every participant is randomly paired with another participant, and each group has two participants.

At the end of these instructions, you are asked several questions to make sure that the instructions are clear.

Contribution Decision

You receive an endowment of 20 tokens. You decide how many of these 20 tokens to contribute to a project (and how many to keep for yourself). Your partner makes the same decision, and s/he can also either contribute tokens to the project or keep tokens for him/herself. You and your partner can choose any number of tokens to contribute between 0 and 20 tokens. Every token that you do not contribute to the project belongs to you and will be paid in Chinese Yuan to you at the end of the experiment.

[36] Note for the reader: These correspond to Stages-1 and Stage-2 in subsection 5.1.

The total investment (G) in the project is the sum of the amounts contributed by you and your partner. If you contribute x tokens and s/he contributes y tokens, then the total investment in the project is $G = x + y$. The project generates a value 1.6 times G, which is shared equally between you and your partner. For instance, if you and your partner each contribute 5 tokens ($x = 5$ and $y = 5$) then $G = 5 + 5 = 10$ tokens. The value of the project is then 1.6 times 10 tokens, or 16 tokens, which are shared equally between you and your partner, i.e., 8 tokens each.

Guess Your Partner's Contribution Decision

Before you make the contribution decision, you are asked to guess how much your partner will contribute to the project. Write down your guess (any number between 0 to 20 tokens) on the Guess Sheet.

You will have a chance to win an additional prize. At the end of the experiment, we will randomly choose one participant whose guess matches his/her partner's actual contribution, and give this participant a prize of 10 Yuan. If nobody guessed correctly, then we will randomly choose one participant whose guess is the closest to the partner's actual contribution, and give this participant a prize of 2 Yuan.

When you complete the guess sheet, the experimenter will collect it. After this, you receive the Decision Sheet. You make your contribution decisions by following the instructions on the Decision Sheet.

How is your income calculated from your contribution decision?

The income of all participants is calculated in the same way. Your income consists of two parts:

(1) The tokens that you keep for yourself (i.e. the income from tokens kept).

(2) The income from the project. The formula for this income is the following

$$1.6\times(\text{sum of all tokens contributed to the project})/2$$
$$= 0.8\times(\text{sum of all tokens contributed to the project}).$$

Therefore, your total income will be calculated by the following formula:

$$(20 - \text{the tokens you contributed to project}) + 0.8\times(\text{sum of all tokens contributed to project}).$$

In order to explain the income calculation consider the following example:

Suppose that you contribute 20 tokens, and your partner contributes 10 tokens. Each of you will receive:

$$0.8(10 + 20) = 0.8 \times 30 = 24 \text{ tokens from the project.}$$

You contribute all your 20 tokens to the project. You will therefore receive 24 tokens in total at the end of the experiment.

Your partner also receives 24 tokens from the project. In addition, s/he receives 10 tokens (the income from tokens kept) because s/he contributed only 10 tokens to the project (thus, 10 tokens remain for him/herself), and s/he receives $24 + 10 = 34$ tokens altogether.

Calculation of your total income in tokens: $(20 - 20) + 0.8 \times (20 + 10) = 24$

Calculation of the total income of your partner in tokens: $(20 - 10) + 0.8 \times (20 + 10) = 34$

Control questions

The following questions are hypothetical and only serve to enhance understand of the income calculations. In these questions, you do not need to consider the prize from correctly guessing your partner's contributions or making the closest guess.

Question 1. Both you and your partner contribute 0 tokens to the project. What is, in tokens,

- your total income?
- your partner's total income?

Question 2. Both you and your partner contribute 20 tokens. What is, in tokens,

- your total income?
- your partner's total income?

Question 3. You contribute 13 tokens. Your partner contributes 8 tokens. What is, in tokens,

- your total income?
- your partner's total income?

Question 4. You contribute 5 tokens. Your partner contributes 11 tokens. What is, in tokens,

- your total income?
- your partner's total income?

Instruction for the first part[37]

In this part, you will be randomly paired with a participant. You will never learn who your partner is.

Please write down your guess of your partner's possible contribution on the Guess Sheet. The Guess Sheet will be collected when you complete it. The remaining instruction for the first part will be then given to you.

[37]Note for the reader: This corresponds to Stages-1 in subsection 5.1. Half of the subjects, the information-advantageous group received this set of instructions. The other half received the instructions for the first part that follow after this set of instructions. Each group of players (the information-advantageous group and the remaining group) were not aware that other subjects may not be receiving identical instructions.

Guess Sheet

What do you believe is the amount that your partner will contribute? Please choose any number between 0 and 20 tokens: _____ tokens.

Instruction for the first part continued...

You will be informed about your partner's guess after both parts of the experiment are complete. However, **your partner doesn't know that you will be informed about his/her guess, and s/he is not informed about your guess**. Please fill in **every** row in the second column. Your payoff-relevant contribution is the amount that you choose corresponding to your partner's actual guess.

For each level of the known guess of your partner about your contribution (see inputs in column) choose your contribution in tokens (any number between 0 and 20):

Decision Sheet

If your partner's guess of your contribution is the following tokens (see inputs in this column)...	... then you contribute the following amount of tokens (any number between 0 and 20):
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

Instruction for the first part[38]

In this part, you will be randomly paired with a participant. You will never learn who your partner is.

Please write down your guess of your partner's possible contribution on the *Guess Sheet*. The *Guess Sheet* will be collected when you complete it. The remaining instructions for the first part will be then given to you.

Guess Sheet

What do you believe is the amount that your partner will contribute?

Please choose any number between 0 and 20 tokens: _____tokens.

[38]Note for the reader: These were the instructions for the first part (corresponds to Stages-1 in subsection 5.1) that were given to the remaining group of players who were not the information-advantageous group (see also previous footnote).

Decision Sheet

What is your contribution to the project?

Please choose any number between 0 and 20 tokens: _____tokens.

Instruction for the second part[39]

In this part, you will be **randomly** paired with another participant (your partner is **different** from that in the first part). You will never learn who your partner is.

Please write down your guess of your partner's possible contribution on the *Guess Sheet*. The *Guess Sheet* will be collected when you complete, and the rest instruction for the first part will be then given to you.

Guess Sheet

What do you believe is the amount that your partner will contribute? Please choose any number between 0 and 20 tokens: _____tokens.

Instruction for the second part continued...

You will be informed about your partner's guess after both parts of the experiment are complete. Your partner knows that you will be informed about his/her guess. And your guess will also be revealed to your partner after both parts are complete. Please fill in every row in the second column. Your payoff-relevant contribution is the amount that you choose corresponding to your partner's actual guess.

Decision Sheet[40]

Post-experimental Questionnaire

1. Age: _____ years old

Gender: (female/male)

Field of study: _____

Degree of study: _____

Year of study: _____

2. Have you participated in similar experiments in the past? (Yes/No)

3. How did you form beliefs about your partner's contribution?

A. You used your own 'desired contribution' (i.e. what you want to contribute) to predict your partner's contribution.

[39]Note for the reader: This corresponds to Stage-2 in subsection 5.1. These are the instructions for the public treatment in our experiment. The private and public treatment were run in a counterbalanced order.

[40]Note for the reader: This decision sheet is the same with the one for the information advantageous subjects in the APR treatment.

B. You used information other than in A to predict your partner's choice. (Please specify)

4. What do you think is your partner's expectation of your contribution in the first part? _____ tokens (any number between 0 and 20).[41]

What do you think is your partner's expectation of your contribution in the second part? _____ tokens (any number between 0 and 20).

10.3. Appendix C: Experimental instructions for the between-subject design

The instructions for the between–subjects design are very similar to the within–subjects design with the following two main differences. First, no strategy method was used to elicit the contribution decisions of the players. Second, in the within–subjects design, the same set of subjects played all treatments in a counterbalanced manner. However, in the between–subjects design, subjects played one of the following three treatments: the *private treatment*, the *asymmetric private treatment* or the *public treatment*. The only difference in the private treatment from the asymmetric private treatment was the absence of the information-advantageous group. Detailed instructions, if required, are available from the authors.

10.4. Appendix D: More on psychological utility functions

Recall, from subsection 3.5, that a player suffers disutility if he thinks he has negatively surprised his partner. Yet, maybe surprisingly, he himself does not suffer disutility from a negative surprise inflicted on him by his partner. And similarly for positive surprises and the intentions behind positive and negative surprises. In this subsection, we rectify this possible omission by including extra terms in the utility functions. We shall see that none of these extra terms changes any of our results and, hence, they were omitted from the rest of the paper. However, their inclusion here helps motivate the other, choice-relevant, terms in the utility functions that were retained in subsection 3.5. Furthermore, we believe that the fuller description of the utility functions given in this subsection helps to better appreciate the nature of psychological utility.

We start with an example that is an analogue of 2 but for first order beliefs of player 1 in the PUB treatment.

Example 3 *: We consider a two-player public goods game. Each player has the initial endowment $y = 2$. Player i contributes $g_i \in [0, 2]$ to the public good, $i = 1, 2$. We consider the public treatment (PUB). Player* 1 *has a first order belief about the contribution, g_2,*

[41] Note for the reader: Subjects were asked Q4 before they were informed of the partner's first order beliefs.

made by player 2 that is given by the probability density $f_1^1(x)$, $x \in [0,2]$. *Player 1 reports a statistic,* θ_2, *about* $f_1^1(x)$, *for example the mean, the median or the mode (or any other statistic) of his privately known belief distribution,* f_1^1. *Player 1 knows that* θ_2 *is communicated to player 2 before player 2 decides on his contribution (in fact,* θ_2 *is made public knowledge). Having sent the signal* θ_2 *to player 2, player 1 updates his belief by using the conditional distribution* $f_1^1(x|\theta_2)$. *In this Example, we shall assume that* θ_2 *is what player 1 regards as the most probable value for* g_2. *For the purposes of this Example, we take the first order belief of player 1 to have the conditional probability density:*

$$f_1^1(x|\theta_2) = \frac{x}{\theta_2},\ x \in [0,\theta_2],\ \theta_2 \in (0,2], \tag{10.16}$$

$$f_1^1(x|\theta_2) = \frac{2-x}{2-\theta_2},\ x \in [\theta_2,2],\ \theta_2 \in [0,2). \tag{10.17}$$

Geometrically, the density (10.16), (10.17) forms the two sides of a triangle with base length 2 and height 1 (so the area under the density is 1, as it should be). The apex of the triangle is at θ_2. *Hence, player 1 believes that player 2 will most probably contribute* $g_2 = \theta_2$. *Suppose, for instance, that* $\theta_2 = 2$. *From (10.16) we get* $f_1^1(x|2) = \frac{x}{2}$, $x \in [0,2]$. *In this case, player 1 believes that player 2 will most probably make the maximum contribution,* $g_2 = 2$. *At the other extreme, suppose that* $\theta_2 = 0$. *From (10.17) we get* $f_1^1(x|0) = 1 - \frac{x}{2}$, $x \in [0,2]$. *Here, player 1 thinks that player 2 will most probably contribute nothing,* $g_2 = 0$. *The cumulative conditional distributions corresponding to (10.16) and (10.17) are, respectively,*

$$F_1^1(x|\theta_2) = \frac{x^2}{2\theta_2},\ x \in [0,\theta_2],\ \theta_2 \in (0,2]. \tag{10.18}$$

$$F_1^1(x|\theta_2) = \frac{2x - \frac{1}{2}x^2 - \theta_2}{2-\theta_2},\ x \in [\theta_2,2],\ \theta_2 \in [0,2). \tag{10.19}$$

A large number (in fact, an infinite number) of unconditional distributions are consistent with (10.16)-(10.19). For example, let player 1's prior distribution of θ_2 *(before he sends the signal containing a realization of* θ_2*) be:*

$$\pi_1^1(\theta_2) = 1 - \frac{1}{2}\theta_2, \theta_2 \in [0,2], \tag{10.20}$$

According to (10.20), player 1 believes that the most probable contribution of player 2 is zero. But many other prior distributions are consistent with (10.16)-(10.19), including:

$$\pi_1^1(\theta_2) = \frac{1}{2}\theta_2, \theta_2 \in [0,2], \tag{10.21}$$

according to which player 1 believes that the most probable contribution of player 2 is all

his endowment. Using

$$f_1^1(x) = \int_{\theta=0}^{\theta=2} f_1^1(x|\theta)\,\pi_1^1(\theta)\,d\theta, \tag{10.22}$$

then (10.20), along with (10.16) and (10.17), imply the unconditional density:

$$f_1^1(0) = 0,\ f_1^1(x) = (\ln 2)\,x - x\ln x,\ x \in (0,2], \tag{10.23}$$

and, hence, the unconditional cumulative distribution:

$$F_1^1(0) = 0,\ F_1^1(x) = \frac{1}{4}x^2 + \frac{1}{2}(\ln 2)\,x^2 - \frac{1}{2}x^2\ln x,\ x \in (0,2]. \tag{10.24}$$

Of course, had we used (10.21) instead of (10.20), in conjunction with (10.16), (10.17) and (10.22), we would have got unconditional distributions different from (10.23) and (10.24).

10.4.1. Psychological utility for the APR treatment

Recall that the psychological utility function of a player 1 was given by (3.12) in subsection 3.5. It is now given by (10.25), below, and the *psychological utility function* of a player 2 in APR2 is now given by (10.26), below that.

$$\begin{aligned} U_1^{APR}(g_1,g_2,\theta_1) &= u_1(g_1,g_2) + \psi_1^S(g_2) + \phi_1^S(g_1,\theta_1) + \psi_1^I(g_2) + \phi_1^I(g_1), & (10.25)\\ U_2^{APR}(g_2,g_1) &= u_2(g_2,g_1) + \psi_2^S(g_1) + \phi_2^S(g_2) + \psi_2^I(g_1) + \phi_2^I(g_2). & (10.26) \end{aligned}$$

Player 1 (who is in APR1) is the informed player, and he receives a signal, θ_1, about what player 2 expects him to contribute. Player 2 (who is in APR2) is the uninformed partner, receives no signal. Hence, the utility of player 1, in (10.25), depends on θ_1 but the utility of player 2, in (10.26), does not depend on a signal.

Note that $\psi_1^S(g_2)$, $\psi_1^I(g_2)$ in (10.25) depend on g_2 but not on g_1. Since player 2 decides on g_2 before he observes g_1, his choice of g_2 cannot be affected by player 1's choice of g_1. Hence, for player 1's decision problem, the two terms $\psi_1^S(g_2)$, $\psi_1^I(g_2)$ do not influence the choice of g_1 (but, of course, they contribute to the utility of player 1). Hence, they were dropped from (3.12) in subsection 3.5 without affecting any of the results. Similar remarks apply to the two functions $\psi_2^S(g_1)$, $\psi_2^I(g_1)$ in (10.26). These four functions are absent from Khalmetski et al. (2015) but we believe that they are important to motivate the other four functions $\phi_1^S(g_1,\theta_1)$, $\phi_1^I(g_1)$, $\phi_2^S(g_2)$, $\phi_2^I(g_2)$ in (3.12) and (3.12) that do affect choices. Let

$$\mu_i \in [0,1],\ \gamma_i \geq 0,\ \delta_i \geq 0, i = 1,2, \tag{10.27}$$

these complement the parameters in (3.14).Consider the function $\psi_1^S(g_2)$ in (10.25). Ex-ante, player 1 expects player 2 to contribute $x \in [0, y]$ with probability density $f_1^1(x)$. Ex-post, player 1 discovers that player 2 has actually contributed $g_2 \in [0, y]$. For $x \in [0, g_2]$, player 1 is pleasantly *surprised*. For $x \in [g_2, y]$, player 1 is *disappointed*. Specifically,

$$\psi_1^S(g_2) = \mu_1 \left\{ \gamma_1 \left[\int_{x=0}^{g_2} (g_2 - x) f_1^1(x) dx \right] - \delta_1 \left[\int_{x=g_2}^{y} (x - g_2) f_1^1(x) dx \right] \right\}. \quad (10.28)$$

If $\psi_1^S(g_2) > 0$, then, on balance, player 1 is pleasantly surprised. Conversely, if $\psi_1^S(g_2) < 0$, then, on balance, player 1 is disappointed. We call $\psi_1^S(g_2)$ *the surprise function for player* 1. Analogously, *the surprise function for player* 2, $\psi_2^S(g_1)$ in (10.26) is defined by

$$\psi_2^S(g_1) = \mu_2 \left\{ \gamma_2 \left[\int_{x=0}^{g_1} (g_1 - x) f_2^1(x) dx \right] - \delta_2 \left[\int_{x=g_1}^{y} (x - g_1) f_2^1(x) dx \right] \right\}. \quad (10.29)$$

Given that player 1 is aware of his own surprise function, $\psi_1^S(g_2)$, it may be reasonable to assume that he attributes a surprise function, $\psi_2^S(g_1)$, to player 2.[42] Assuming that player 1 has a degree of empathy for player 2, it is reasonable to assume that player 1 gains utility from positively surprising player 2 but suffers a utility loss by negatively surprising player 2. This was formalized by the function $\phi_1^S(g_1, \theta_1)$ in (3.12) and (3.15) of subsection 3.5 and retained in (10.25) above. Analogously for $\phi_2^S(g_2)$ in (3.13) and (3.16) of subsection 3.5 and retained in (10.26) above. Recall that $\phi_2^S(g_2)$ does not depend on a signal. This is because, since player 2 is the uninformed player, he does not receive a signal to condition on.

Now, consider the function $\psi_1^I(g_2)$ in (10.25) above, and (10.30) below.

$$\psi_1^I(g_2) = (1 - \mu_1) \left\{ \gamma_1 \left[\int_{x=0}^{g_2} (g_2 - x) f_1^3(x) dx \right] - \delta_1 \left[\int_{x=g_2}^{y} (x - g_2) f_1^3(x) dx \right] \right\}. \quad (10.30)$$

Recall that f_1^3 represents the beliefs of player 1 about the second order beliefs of player 2, f_2^2, which in turn are beliefs of player 2 about player $1's$ first order beliefs f_1^1. In (10.30), player 1 believes, with probability density $f_1^3(x)$, that player 2 thinks that player 1 expects player 2 to contribute $x \in [0, y]$. For $x \in [0, g_2]$, player 1 gains an expected utility $(1 - \mu_1) \gamma_1 \int_{x=0}^{g_2} (g_2 - x) f_1^3(x) dx$. For $x \in [g_2, y]$, player 1's expected utility is decreased by $(1 - \mu_1) \delta_1 \int_{x=g_2}^{y} (x - g_2) f_1^3(x) dx$. As an illustration, suppose $\psi_1^S(g_2) < 0$, so player 1 suffers *negative surprise*. This pain to player 1 would be ameliorated if player 1 believed that, when player 2 chose g_2, then player 2 thought that he would be delivering a *positive surprise* to player 1 (when, in fact, player 2 delivered a negative surprise to player 1).[43]

[42] This can be formalized using evidential reasoning. See, for example, al-Nowaihi and Dhami (2015).

[43] This makes sense because we do not require consistency of action and beliefs, see Section 3.4 above.

In this case $\psi_1^I(g_2) > 0$. On the other hand, this pain to player 1 would be increased if player 1 believed that, when player 2 chose g_2, then player 2 thought that he would be delivering a *negative surprise* to player 1. In this case $\psi_1^I(g_2) < 0$.[44] Thus, we call $\psi_1^I(g_2)$ the *intentional surprise function for player* 1. Analogously, $\psi_2^I(g_1)$, in (10.26) above, and (10.31) below, we call *the intentional surprise function for player* 2.

$$\psi_2^I(g_1) = (1-\mu_2)\left\{\gamma_2\left[\int_{x=0}^{g_1}(g_1-x)\,f_2^3(x)\,dx\right] - \delta_2\left[\int_{x=g_1}^{y}(x-g_1)\,f_2^3(x)\,dx\right]\right\}. \quad (10.31)$$

We now give an argument to motivate $\phi_1^I(g_1)$ in (3.12) and (3.17) of subsection 3.5 and retained in (10.25), above, that is similar to the argument we gave to motivate $\phi_1^S(g_1)$. Given that player 1 is aware of his own intentional surprise function, $\psi_1^I(g_2)$, it may be reasonable to assume that he attributes an intentional surprise function, $\psi_2^I(g_1)$, to player 2. Assuming that player 1 has a degree of empathy for player 2, it is reasonable to assume that player 1 gains utility from believing that player 2 thinks that player 1 intended to positively surprise him but suffers a utility loss from believing that player 2 thinks that player 1 intended to negatively surprising him. This is formalized by the function $\phi_1^I(g_1)$. Analogously for $\phi_2^I(g_2)$ in (3.13) and (3.18) of subsection 3.5 and retained in (10.26) above.

10.4.2. Psychological utility for the PUB treatment

Recall that in PUB each player, i, receives a signal, θ_i, about the contribution, g_i, that his partner, player $-i$, expects him (player i) to make. Furthermore, each player i knows that his partner, player $-i$, has received that signal and this is public knowledge. If follows that the densities that enter the psychological utility function for player i in PUB are conditional on θ_i. Hence, the psychological utility function of player i in PUB is given by:

$$U_i^{PUB}(g_i, g_{-i}, \theta_i, \theta_{-i}) = u_i(g_i, g_{-i}) + \psi_i^S(g_{-i}, \theta_{-i}) + \phi_i^S(g_i, \theta_i) + \psi_i^I(g_{-i}, \theta_{-i}) + \phi_i^I(g_i, \theta_i), \quad (10.32)$$

where the functions $\psi_i^S(g_{-i}, \theta_{-i})$ and $\psi_i^I(g_{-i}, \theta_{-i})$ are given by:

$$\psi_i^S(g_{-i}, \theta_{-i}) = \mu_i\left\{\gamma_i\left[\int_{x=0}^{g_{-i}}(g_{-i}-x)\,f_i^1(x|\theta_{-i})\,dx\right] - \delta_i\left[\int_{x=g-i}^{y}(x-g_{-i})\,f_i^1(x|\theta_{-i})\,dx\right]\right\}, \quad (10.33)$$

$$\psi_i^I(g_{-i}, \theta_{-i}) = (1-\mu_i)\left\{\gamma_i\left[\int_{x=0}^{g_{-i}}(g_{-i}-x)\,f_i^3(x|\theta_{-i})\,dx\right] - \delta_i\left[\int_{x=g_{-i}}^{y}(x-g_{-i})\,f_i^3(x|\theta_{-i})\,dx\right]\right\}, \quad (10.34)$$

[44] Suppose you stepped on my toe. This is, of course, physically painful to me. Furthermore, suppose that I thought that your action was deliberate rather than accidental. Then, in addition to the physical pain, I would also experience a psychological pain.

and the parameters are as in (10.27) above.

The interpretation of (10.32), (10.33) and (10.34) is the same as (10.25) to (10.31) except for the introduction of the conditioning on θ_i, θ_{-i}.