Essays on Psychological Game Theory and Ambiguity

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Mengxing Wei
Department of Economics
University of Leicester

July 2017

To my parents Weiran Wang, Jinghu Wei, and my fiancé Weijie Chi.

Abstract

This thesis mainly focuses on two themes, psychological game theory and quantum decision theory. Chapter 2 and Chapter 3 study how emotions and other-regarding preferences affect classical results in game theory. Chapter 4 tests the quantum decision theory model of the Ellsberg paradox that has been developed by al-Nowaihi and Dhami (2017).

Chapter 2 models guilt-aversion/surprise-seeking, and the attribution of intentions behind these emotions in a one-shot public goods game. Using the induced beliefs method in both within-subjects design (strategy method) and between-subjects design, the experimental results show that guilt-aversion is predominant relative to surprise-seeking, and the attribution of intentions behind these emotions are important.

Chapter 3 compares three main competing explanations for the choice of effort by workers in a gift exchange game - classical reciprocity (Akerlofs action-based formulation, Malmendier and Schmidt (2017) formulation) and belief-based reciprocity (psychological game theory). Experimental results show that all models explain well about the workers choices of efforts, and psychological game theory can predict their emotions of guilt. However, Akerlofs model is the best in terms of parsimony and fit.

Chapter 4 experimentally tests the *matching probabilities* for the Ellsberg paradox, which is based on a parameter-free theoretical derivation using quantum probabilities rather than Kolmogorov probabilities (al-Nowaihi and Dhami, 2017). The experimental results are consistent with the quantum model, and subjects are ambiguity seeking for the low probabilities but ambiguity averse for the medium and high probabilities.

Acknowledgements

I would like to thank my supervisors and co-authors - Professor Sanjit Dhami and Professor Ali al-Nowaihi - for their guidance and help in my PhD study. I also thank for their patience, motivation, and immense knowledge. Without them, I cannot get improved so much. It is fortunate for me to learn important attitudes from them: the conscientiousness and dedication in doing research, and the optimism and courage in dealing with difficulties. I also appreciate the way they supervised me: pointing out my mistakes and problems when I did badly, while encouraging me when I did well. Now I look back, and I can realize my progress under their supervision (although there is still a long way to go in future). I would like to thank another supervisor of mine - Dr Subir Bose - for his continuous support.

During my PhD study, I have benefited a lot from discussions with Professor Stephen Hall, Dr Jesse Matheson, Dr David Rojo Arjona and Dr Cheng Zhou. I also thank Professor Jianbiao Li for providing me with opportunities to run experiments in China.

I am grateful to my friends, Xin, Guangyi, Jingyi, Xing, Haige, Livia, Emma, Taha, Junaid, Temo, Sam and all others, for their help in my PhD life.

Generous financial support from the Department of Economics and College of Social Sciences, Arts, and Humanities at the University of Leicester is gratefully acknowledged.

My parents, who are the mentors of my life, always support me mentally and financially. Without their encouragement, I could not have been so determined to pursue PhD in the UK. Every time I was faced with obstacles, they directed me to a bright destination. When I achieved even little, they were proud and even happier than me. I just want to say how lucky I am to be their child. My fiancé, Weijie Chi, makes my life more colourful and delightful. Especially these years far away from home, his love was the sunshine to support me to complete this thesis.

Declaration

All chapters of this thesis are joint work with my PhD supervisors, Professor Sanjit Dhami and Professor Ali al-Nowaihi.
Chapter 2 has appeared under the title ”Public goods games and psychological utility: Theory and evidence” in the 2016 University of Leicester Paper Series in Economics. Versions of Chapter 2 were also presented at the following conferences:
- October 2016: Preference, Personality and Moral Behavior, Bergen, Norway.
- September 2016: XXXI Jornadas de Economa Industrial, Palma de Mallorca, Spain.
- September 2015: The 8th International Symposium on Corporate Governance, Tianjin, China.

Chapter 4 has appeared under the title ”Can quantum decision theory explain the Ellsberg paradox?” in the 2017 University of Leicester Paper Series in Economics.

Mengxing Wei
PhD Research Student
Astley Clarke Building PhD Center
Department of Economics
University of Leicester
LE1 7RH
+44 (0) 7544128185
mw323@le.ac.uk
maxine.wmx@gmail.com

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Beliefs shape motivation, and belief-dependent motivations are important in strategic decision making (Geanakoplos, et al., 1989; Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Battigalli and Dufwenberg, 2009), such as guilt and reciprocity. Guilt-aversion is one of the central determinants of human economic behaviour. In psychological game theory, people may feel guilty in proportion to the degree to which they do not live up to others' expectations. Theoretical and experimental studies corroborate the important effects of guilt-aversion (Battigalli and Dufwenberg, 2007, 2009; Dufwenberg, et al., 2011; Khalmetski, et al., 2015). Reciprocity is another important human trait. People are reciprocal if they punish others' unkindness and reward others' kindness. A great deal of experimental evidence shows that reciprocity plays a vital role in strategic interactions (Berg et al., 1995; Fehr et al., 1993; Camerer and Thaler, 1995; Falk et al., 2008).

Chapter 2 investigates the effects of guilt-aversion/surprise-seeking, and the attribution of intentions behind these emotions in decision making. The previous studies using dictator game experiments (Khalmetski, et al., 2015; Ellingsen et al., 2010) may not be robust to extend to strategic environments. Thus, Chapter 2 theoretically and experimentally tests these belief-dependent effects in public goods games. To eliminate the *false consensus effect*, the experiments follow the induced beliefs method in Ellingsen et al. (2010) and Khalmetski, et al. (2015), which is incorporated in both the within-subjects design (strategy method) and between-subjects design. The within-subjects experiments show that the majority of the subjects are relatively guilt-averse, and the attribution of intentions behind these emotions are important in their choices of contributions.

Chapter 3 studies reciprocity in the gift exchange game, where the firm gives a wage offer, and the worker chooses an effort level. The action-based reciprocity requires the worker to respond positively to a gift (wage paid by the firm), irrespective of any other considerations. Akerlof's formulation (1982) and Malmendier and Schmidt (2017) formulation are utilized to model the classical reciprocity. In contrast, conditionally reciprocal players also consider the intentions of the other players. This chapter models belief-based reciprocity in its sequential version in Dufwenberg and Kirchsteiger (2004) and (simple) guilt in Battigalli and Dufwenberg (2007). Experiments are designed to test and compare the three competing models. Experimental results are mostly in accord with the above theories, and wage and beliefs influence the worker's choices of efforts. However, only the psychological game theory model correctly predicts the effects of beliefs; and the positive correlation between the worker's second order beliefs and efforts (significant in the domain of negative reciprocity) provides evidence of the important effect from guilt. Model selection tests picked the Akerlof's action-based formulation as the best in the sense of parsimony and fit.

Chapter 4 explores if quantum decision theory can explain better people's decision making in situations of ambiguity, and it is the first experimental test of the quantum decision model of al-Nowaihi and Dhami (2017). The predictions of al-Nowaihi and Dhami (2017) are in good agreement with the pre-existing data of Dimmock et al. (2015). However, Dimmock et al. (2015) used a nonincentive compatible mechanism. By contrast, Chapter 4 designed and implemented an incentive compatible mechanism in the context of the Ellsberg paradox (Fox and Tversky, 1995) to find the matching probabilities of the unknown urn under the three representative probabilities (0.1, 0.5, and 0.9) of the known urn. The predicted matching probabilities are consistent with those observed in our experiments, and subjects are ambiguity seeking for low probabilities but ambiguity averse for medium and high probabilities. The experiments also investigated the robustness of the quantum decision model to the introduction of various demographic characteristics, such as gender, training in statistics and economics, attainment in education. The demographic characteristics have little or no influence on the subjects' attitudes to ambiguity, however, those with prior training in statistics exhibit results different from those who do not have such training.

These three chapters may also be germane for future research. For instance, how do beliefs and actions in the information set influence guilt in more complicated scenarios such as repeated games? How do more general utility functions and consideration of norms affect the results of Chapter 3? Would the quantum model perform equally well in other situations of ambiguity? Further research is needed to answer the above questions.

Chapter 2

Public goods games and psychological utility: Theory and evidence.

Sanjit Dhami, Mengxing Wei, Ali al-Nowaihi

Abstract

We consider a public goods game which incorporates guilt-aversion/surprise-seeking and the attribution of intentions behind these emotions (Battigalli and Dufwenberg, 2007; Khalmetski et al., 2015). We implement the induced beliefs method (Ellingsen et al., 2010) and a within-subjects design using the strategy method. Previous studies mainly use dictator games - whose results may not be robust to adding strategic components. We find that guilt-aversion is far more important than surprise-seeking; and that the attribution of intentions behind guilt-aversion/surprise-seeking is important. Our between-subjects analysis confirms the results of the within-subjects design.

2.1 Introduction

In classical game theory, the utility functions of the players map the set of strategy profiles into the set of payoffs. We shall refer to classical utility 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 examples illustrate how the feelings of *surprise* and *guilt* may directly impart disutility.

Example 2.1. *: John frequently visits cities A and B, and he typically uses a taxi to get around. In city A, the taxi driver expects no tips, 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 *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; 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.

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 2.1 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 2.1, 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 2.1, 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]

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

[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).

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 2.1, 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.

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

[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.

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 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-seeking but a between-subjects design for the role of

[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).

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? The answer has an important bearing on the use of psychological game theory in economics.

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.[7] 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 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.

The plan of the paper is as follows. Section 2.2 outlines the basic model of public goods. Section 2.3 considers the implications of guilt-aversion and surprise-seeking in a two-player public goods game; and the attribution of intentions behind these. Section 2.4 gives the theoretical predictions of our model. Section 2.5 describes our within-subjects experimental design and Section 2.6 gives our experimental results.

[7]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.

Section 2.7 reconsiders the empirical results in a between-subjects design. Section 2.7.4 examines the determinants of contributions. Section 2.8 concludes. Appendix A.1 contains the proofs. Appendices A.2 and A.3 contain, respectively, the instructions for the within-subjects and the between-subjects designs. Appendix A.4 provides further discussion of the psychological utility functions.

2.2 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.[8] 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, $u_i(g_i, g_{-i}) = v_i(c_i) + r_i(g_i + g_{-i})$, where $r_i > 0$, and v_i is a strictly increasing and strictly concave utility function of private consumption, c_i, so $v_i' > 0$, $v_i'' < 0$[9]. The budget constraint is given by $c_i + g_i = y_i$. Substituting the constraint into the utility function, the utility function becomes

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

The parameter r_i 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 $r_i < v_i'(y)$, i.e., the net return to an individual from a unit of contributions is negative. Since $v_i'' \leq 0$, thus,

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

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

Proposition 2.1. *: 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* $G^n = 0$.

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

[9]We need the property of strict concavity of v_i to derive the unique optimal solution, however, in the experiments we use the linear form for simplicity.

To distinguish the ordinary utility (2.1) from the psychological utility, to be introduced in Section 2.3, immediately below, we shall refer to (2.1) as the *material utility.*

2.3 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.

2.3.1 Levels of beliefs

Let us 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. Throughout our paper, the 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

[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.

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 2.3.2.2, below, we shall introduce conditional fourth order beliefs.

2.3.2 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 2.7. We now discuss the other two treatments that are common to the within-subjects and the between-subjects design.

2.3.2.1 APR treatment

In APR, subjects are divided into two equal groups: APR1 and APR2. Every subject in APR1 is 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 and a subscript 2 to denote a player in APR2. 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.

According to our theory, 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

[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. Furthermore, in order to distinguish higher order beliefs we will need signals of higher order beliefs. It can be shown that the signal θ_i, as defined above, allows us to distinguish beliefs up to order four only.

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 beliefs unconditional on the signal. 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.

2.3.2.2 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.

2.3.3 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[12]

[12]It might appear natural to write Assumption A3 for $k = 1, 2, 3, 4$. However, we only need the cases $k = 2, 4$.

$$\theta_i' > \theta_i \Rightarrow F_i^k(x|\theta_i') < F_i^k(x|\theta_i)$$

$$\text{for all } x \in (0,y) \text{ and } \theta_i', \theta_i \in (0,y), i = 1,2, k = 2,4.$$

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

$$\frac{\partial F_i^k(x|\theta_i)}{\partial \theta_i} < 0 \text{ for all } x \in (0,y) \text{ and all } \theta_i \in (0,y),\ i = 1,2,\ k = 2,4. \tag{2.3}$$

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 2.2. *: 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, 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 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:*

$$f_1^2(x|\theta_1) = \frac{x}{\theta_1},\ x \in [0, \theta_1],\ \theta_1 \in (0, 2], \tag{2.4}$$

$$f_1^2(x|\theta_1) = \frac{2-x}{2-\theta_1},\ x \in [\theta_1, 2],\ \theta_1 \in [0, 2). \tag{2.5}$$

Geometrically, the density (2.4), (2.5) 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* $\theta_1 = 2$*. From (2.4) we get* $f_1^2(x|2) = \frac{x}{2}$*,* $x \in [0, \theta_1]$*. In this case, player* 1 *thinks that player* 2 *believes that player* 1 *will most probably make the maximum contribution,* $g_1 = 2$*. At the other extreme, suppose player* 1 *receives* $\theta_1 = 0$*. From (2.5) we get* $f_1^2(x|0) = 1 - \frac{x}{2}$*,* $x \in [\theta_1, 2]$*. Here, player* 1 *thinks that player* 2 *believes that player* 1 *will most probably contribute nothing,* $g_1 = 0$*. The cumulative conditional distributions corresponding to (2.4) and (2.5) are, respectively,*

$$F_1^2(x|\theta_1) = \frac{x^2}{2\theta_1},\ x \in [0, \theta_1],\ \theta_1 \in (0, 2]. \tag{2.6}$$

$$F_1^2(x|\theta_1) = \frac{2x - \frac{1}{2}x^2 - \theta_1}{2-\theta_1},\ x \in [\theta_1, 2],\ \theta_1 \in [0, 2). \tag{2.7}$$

From (2.6) and (2.7), it is straightforward to show:

$$\frac{\partial F_1^2(x|\theta_1)}{\partial \theta_1} < 0, x \in (0, 2), \theta_1 \in (0, 2), \tag{2.8}$$

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 (2.4)-(2.7). For example, let player 1*'s prior distribution of* θ_1 *(before he received the signal containing a realization of* θ_1*) be:*

$$\pi_1^2(\theta_1) = 1 - \frac{1}{2}\theta_1, \theta_1 \in [0, 2], \tag{2.9}$$

According to (2.9), 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 (2.4)-(2.7), including:*

$$\pi_1^2(\theta_1) = \frac{1}{2}\theta_1, \theta_1 \in [0,2], \tag{2.10}$$

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

$$f_1^2(x) = \int_{\theta=0}^{\theta=2} f_1^2(x|\theta)\, \pi_1^2(\theta)\, d\theta, \tag{2.11}$$

then (2.9), along with (2.4) and (2.5), imply the unconditional density:

$$f_1^2(0) = 0,\ f_1^2(x) = (\ln 2)\, x - x \ln x,\ x \in (0,2], \tag{2.12}$$

and, hence, the unconditional cumulative distribution:

$$F_1^2(0) = 0,\ F_1^2(x) = \frac{1}{4}x^2 + \frac{1}{2}(\ln 2)\, x^2 - \frac{1}{2}x^2 \ln x,\ x \in (0,2]. \tag{2.13}$$

Of course, had we used (2.10) instead of (2.9), in conjunction with (2.4), (2.5) and (2.11), we would have got unconditional distributions different from (2.12) and (2.13).

2.3.4 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).

2.3.5 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 a utility function for one group with the analogous term in the other two groups.

2.3.5.1 Psychological utility for the APR treatment

The *psychological utility function* of a player 1 in group APR1 is given by (2.14), below, and the *psychological utility function* of a player 2 in group APR2 is given by (2.15), below.

$$\begin{aligned} U_1^{APR}(g_1, g_2, \theta_1) &= u_1(g_1, g_2) + \phi_1^S(g_1, \theta_1) + \phi_1^I(g_1), & (2.14) \\ U_2^{APR}(g_2, g_1) &= u_2(g_2, g_1) + \phi_2^S(g_2) + \phi_2^I(g_2). & (2.15) \end{aligned}$$

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 (2.14), depends on θ_1 but the utility of player 2, in (2.15), does not depend on any signal.

Recall that $u_1(g_1, g_2)$ in (2.14) and $u_2(g_2, g_1)$ in (2.15) are the same utility functions as in the classical public goods game, (2.1). We now explain the various terms in (2.14) and (2.15) and give precise specifications for them. Let

$$\nu_i \in [0,1], \alpha_i \geq 0, \beta_i \geq 0, i = 1, 2. \tag{2.16}$$

Assuming that player 1 has a degree of empathy for player 2, it is possible 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 (2.14) above, and (2.17) below.

$$\begin{aligned} \phi_1^S(g_1, \theta_1) = \nu_1 \Bigg\{ & \alpha_1 \left[\int_{x=0}^{g_1} (g_1 - x) f_1^2(x|\theta_1)\, dx \right] \\ & - \beta_1 \left[\int_{x=g_1}^{y} (x - g_1) f_1^2(x|\theta_1)\, dx \right] \Bigg\}. \end{aligned} \tag{2.17}$$

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*. He aims to pleasantly surprise player 2. 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*.[13] Thus $\phi_1^S(g_1, \theta_1)$ is called *the simple surprise function for player* 1.[14] Analogously, $\phi_2^S(g_2)$, in (2.15) above, and (2.18) 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.

$$\phi_2^S(g_2) = \nu_2 \left\{ \alpha_2 \left[\int_{x=0}^{g_2} (g_2 - x) f_2^2(x) \, dx \right] - \beta_2 \left[\int_{x=g_2}^{y} (x - g_2) f_2^2(x) \, dx \right] \right\}. \tag{2.18}$$

Assuming that player 1 has a degree of empathy for player 2, it is possible 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.[15] This is formalized by the function $\phi_1^I(g_1)$ in (2.14) above, and (2.19) below.[16]

$$\begin{aligned} \phi_1^I(g_1) = (1 - \nu_1) \Bigg\{ & \alpha_1 \left[\int_{x=0}^{g_1} (g_1 - x) f_1^4(x) \, dx \right] \\ & - \beta_1 \left[\int_{x=g_1}^{y} (x - g_1) f_1^4(x) \, dx \right] \Bigg\}. \end{aligned} \tag{2.19}$$

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 $x \in [0, y]$. Thus, $\phi_1^I(g_1)$ is called *the attribution of intentions function for player* 1.

[13] Thus, a player suffers disutility if he thinks he has negatively surprised his partner (for the papers on *disappointment*, see Bell (1985), Loomes and Sugden (1986), etc.). 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 Dhami et al. (2016). 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.

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

[15] 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.

[16] Following Khalmetski et al. (2015), α_1 and β_1 in (2.17) and (2.19) are identical.

Remark 2.1. *: In writing $\phi_1^I(g_1)$, we really should have used the conditional distributions $f_2^3(x \mid g_1)$ and $f_1^4(x \mid g_1)$. After all, player 2 can observe the contribution, g_1, of player 1, and update his third order beliefs, f_2^3, which in turn introduces conditional fourth order beliefs of player 1, f_1^4. There are two reasons we continue to use unconditional distributions for brevity. First, a purely practical consideration is that it does not change any results in our paper, but introduces extra terms and assumptions.*[17] *The second point is a conceptual one. Guilt is an "internal psychological mechanism" that plays an essential role in norm maintenance in societies, independent of external sanctions or higher order beliefs (Bicchieri, 2006; Elster, 1989). However, the attribution of intentions functions rely on what others think of one's actions. In our experiments, players engage in the APR and the PUB treatments first, and at the end of the experiment, the experimenter uses the contributions of the players, anonymously, and purely for determining the payoffs of the players. At this point, the experiment ends. In order to condition f_2^3 and f_1^4 on g_1, we would require the additional assumption that players care about the emotions of other players once they leave the experiment. In contrast, simple guilt–aversion and simple surprise–seeking, that are likely to rely on internal psychological mechanisms, are not subject to this concern. Perhaps, a different experimental design might be needed to rigorously test this idea, which we prefer to leave for future work.*

Analogously, $\phi_2^I(g_2)$, in (2.15) and (2.20), is *the attribution of intentions function for player* 2.

$$\begin{aligned}\phi_2^I(g_2) = (1-\nu_2)\Bigg\{\alpha_2\left[\int_{x=0}^{g_2}(g_2-x)\,f_2^4(x)\,dx\right] \\ -\beta_2\left[\int_{x=g_2}^{y}(x-g_2)\,f_2^4(x)\,dx\right]\Bigg\}.\end{aligned} \tag{2.20}$$

Finally, note that for the surprise function for player 1, $\phi_1^S(g_1,\theta_1)$ in (2.14) and (2.17), 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, $\phi_1^I(g_1)$ in (2.14) and (2.19), 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.

[17]There is no effect on our Propositions 2.2, 2.3, 2.4, 2.5, and 2.6.

2.3.5.2 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:

$$U_i^{PUB}(g_i, g_{-i}, \theta_i) = u_i(g_i, g_{-i}) + \phi_i^S(g_i, \theta_i) + \phi_i^I(g_i, \theta_i), \tag{2.21}$$

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

$$\begin{aligned}\phi_i^S(g_i, \theta_i) = \nu_i \Bigg\{ & \alpha_i \left[\int_{x=0}^{g_i} (g_i - x) f_i^2(x|\theta_i)\, dx \right] \\ & - \beta_i \left[\int_{x=g_i}^{y} (x - g_i) f_i^2(x|\theta_i)\, dx \right] \Bigg\},\end{aligned} \tag{2.22}$$

$$\begin{aligned}\phi_i^I(g_i, \theta_i) = (1-\nu_i) \Bigg\{ & \alpha_i \left[\int_{x=0}^{g_i} (g_i - x) f_i^4(x|\theta_i)\, dx \right] \\ & - \beta_i \left[\int_{x=g_i}^{y} (x - g_i) f_i^4(x|\theta_i)\, dx \right] \Bigg\},\end{aligned} \tag{2.23}$$

and the parameters are as in (2.16) above.

The interpretation of (2.21), (2.22) and (2.23) is the same as (2.14) to (2.20) 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 $f_i^4(x|\theta_i)$.

From (2.14) and (2.21), for $i = 1$ in the latter, we see that the only terms in which $U_1^{APR}(g_1, g_2, \theta_1)$ and $U_1^{PUB}(g_1, g_2, \theta_1)$ differ are $\phi_1^I(g_1)$ and $\phi_1^I(g_1, \theta_1)$. From (2.19) and (2.23), for $i = 1$ in the latter, we see that these terms disappear if, and only if, $\nu_1 = 1$. The latter is the case if, and only if, intentions are unimportant. This leads to a player's contribution under APR1 being the same under PUB (see Proposition 2.6, below). Thus, if the contribution of a player under APR1 is different from the contribution of that same player under PUB, then intentions have to be important for that player.

2.3.6 Psychological equilibria

Recall that in Section 2.3.4, as in Khalmetski et al. (2015) and Ellingsen et al. (2010), we assumed that we do not force consistency of beliefs and actions. Also recall the description of the APR and the PUB treatments (Subsection 2.3.2). 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 (2.14) 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 (2.15) 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 (2.21), 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 (2.21), 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* $\widetilde{g}_1$*).*

2.4 Theoretical Predictions

In this section we derive the theoretical predictions of our model (all proofs are in the Appendix A.1). 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 2.2. *: A psychological equilibrium exists, and is in dominant actions.*

[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.

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 2.3. *: If guilt-aversion is more 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 (2.17), (2.18)). 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 2.4. *(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* $\frac{\partial \widehat{g}_1}{\partial \beta_1} \geq 0$,
(ii) $\frac{\partial \widehat{g}_1}{\partial \alpha_1} > 0$ *and* $\frac{\partial \widehat{g}_1}{\partial \beta_1} > 0$ *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) $\frac{\partial \widehat{g}_2}{\partial \alpha_2} > 0$ *and* $\frac{\partial \widehat{g}_2}{\partial \beta_2} > 0$ *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) $\frac{\partial g_i^*}{\partial \alpha_i} > 0$ *and* $\frac{\partial g_i^*}{\partial \beta_i} > 0$ *for* $\nu_i > 0$ *and* $F_i^2(g_i^*|\theta_i) < 1$ *or* $\nu_1 < 1$ *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 2.5. *: (Comparative statics with respect to* θ_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: For $\nu_1 = 0$, $\frac{\partial \widehat{g}_1}{\partial \theta_1} = 0$, *and for* $\nu_1 > 0$, $\frac{\partial \widehat{g}_1}{\partial \theta_1} \gtreqqless 0 \Leftrightarrow \alpha_1 \lesseqqgtr \beta_1$,
(b) Players in the PUB treatment: $\frac{\partial g_i^*}{\partial \theta_i} \gtreqqless 0 \Leftrightarrow \alpha_i \lesseqqgtr \beta_i$, $i = 1, 2$.

Proposition 2.5 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 Khalmestski et al. (2015), that we describe in Section 2.5 below.

Proposition 2.6. *: Suppose* $\widehat{g}_1, g_1^* \in [0, y]$ *and* $\alpha_1 \leq \beta_1$*. If intentions are unimportant (*$\nu_1 = 1$*), then* $\widehat{g}_1 = g_1^*$*.*

According to Proposition 2.6, 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 2.6.2, that this 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 important (recall Subsection 2.3.5).

Remark 2.2. *: Proposition 2.6 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. However, if* $\widehat{g}_1 = g_1^*$*, then we cannot infer that intentions are unimportant for that player. This remark will play an important role in interpreting our experimental findings.*

If intentions are unimportant then (and only then), the choice relevant terms in the utility function are the same whether the player is under the PUB treatment or under the APR1 treatment; compare (2.14), (2.19) and $\nu_1 = 1$, on the one hand, with (2.21) and (2.23) for $i = 1$ and $\nu_1 = 1$ on the other. Assuming that guilt-aversion is more important than surprise-seeking, then the optimum (in both cases) is unique (Proposition 2.3). 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; compare (2.15), (2.20) and $\nu_2 = 1$, on the one hand, with (2.21), (2.23) for $i = 2$ and $\nu_2 = 1$ on the other. So, we cannot say anything, in general, about the level of contribution under APR2 and PUB, even for a player under identical information conditions in stage 1 of our experiments (see subsections 2.5.1 and 2.5.2). It all depends on the specifics of the probability distributions.

2.5 Within-subjects experimental design

We consider two treatments in our within-subjects design: The *asymmetric private treatment* (APR) and the *public treatment* (PUB).[19]

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 2.5. Also 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). In order to check for the comparability with several earlier papers we also employ the between-subjects design that is described in Section 2.7; however, it cannot be employed to test Proposition 2.5.

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.

2.5.1 Asymmetric private treatment (APR)

The APR treatment, which is described in detail in Appendix A.2, has the following stages.

[19] 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 2.7. 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.

Stage-1: Subjects are initially asked to guess their partner's possible contribution to the public good on a *Guess Sheet* that allows guesses from zero to 20 tokens.[20]

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.

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. This is akin to the *strategy method*.

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 belief, 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 of the public goods game is implemented.

2.5.2 Public treatment

In the public treatment, PUB, the first stage is identical to the APR treatment described in Subsection 2.5.1. 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. This is the essence of the strategy method. Each player is told: "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."

[20]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.

The provision of this information distinguishes the PUB treatment from the APR treatment in the following respect. Each player i, $i = 1, 2$, can condition on both signals, θ_i and θ_{-i} (and not just one of the players and one of the signals), and both players know this.

2.6 Results and discussion of the within-subjects design

2.6.1 Testing Proposition 2.5: Surprise-seeking or guilt-aversion?

Proposition 2.5 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 (such an individual is indexed with the subscript 1).[21] 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 are formally equivalent to 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).

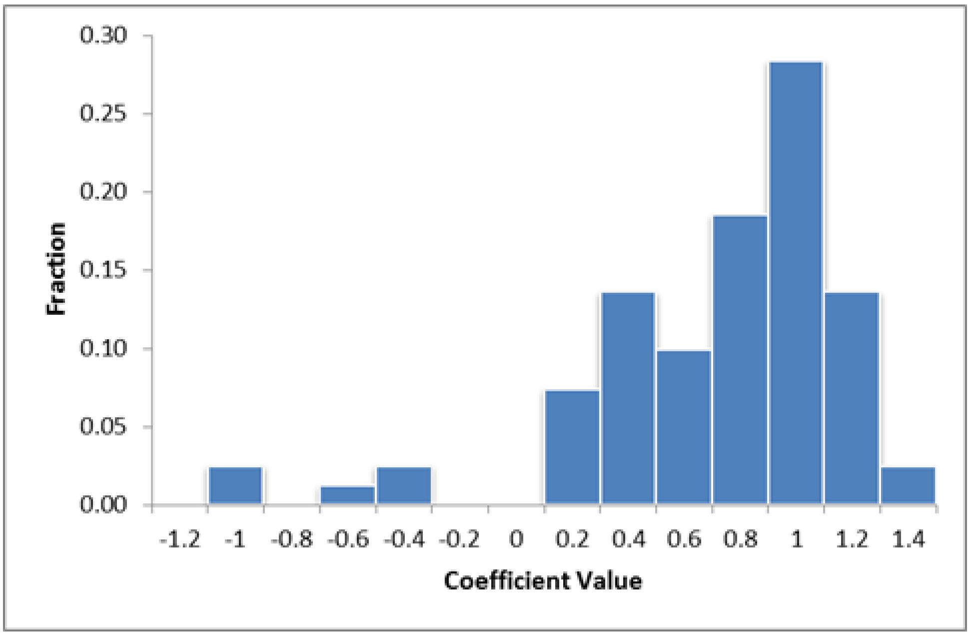
The resulting distribution of the regression coefficients that are significant at the 5% level is shown in Figure 2.1. In Figure 2.1, 5% of the subjects exhibit negative coefficients, and the remaining 95% of the subjects exhibit positive coefficients; Proposition 2.5 predicts that the former are relatively 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

[21] 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 A.2.

Figure 2.1: The distribution of regression coefficients of contributions on guesses (second order beliefs) that are significant at the 5% level in a within-subjects regression.



to our results. In their analogue of Figure 2.1, more than 70% of the coefficients are distributed to the right of zero (compare this to 95% positive coefficients in Figure 2.1). 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. The Spearman correlation coefficient is 0.47 and 0.43, respectively, in the APR1 and PUB treatments; $p = 0.000$ 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 earlier result on the importance of guilt-aversion that used neither within-subjects nor the induced beliefs method can be recovered at the individual and at the aggregate level (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 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.

2.6.2 Testing Proposition 2.6: The importance of intentions

Proposition 2.6 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.[22] We now test this prediction.

Table 2.1: Average first order belief (FOB) and the average contributions.

	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)

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.

Table 2.1 shows the summary statistics of the first stage guesses of the players (the first order beliefs, denoted by FOB, recall Subsection 2.3.1) and the contributions of players in both treatments (APR and PUB) in Experiments 1 and 2; recall Section 2.5. From Table 2.1, we see that contributions range from 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 2.1, 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 in the APR treatment and in the PUB treatment, we have 21 conditional contribution decisions in each treatment. A two-sided Mann-Whitney U test showed that at the 5% significance level, 17 out of 54 subjects in Experiment 1 made significantly different contribution decisions in the two treatments.[23] In Experiment 2, 16 out of 57 subjects made significantly different contri-

[22]Recall Remark 2.2.

[23]The terms ‘Experiment 1’ and ‘Experiment 2’ are defined in Section 2.5.

bution decisions. To further understand the direction of the difference, we compared the mean difference in the two treatments in Table 2.2 using a two-sided t test for equality of means.

Table 2.2: Proportions of Intentional Surprise.

	Higher in PUB	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%

Note: The proportions reported in the table comprise the subset that is significant at 5% in which the mean contributions under PUB are relatively higher. In the third column the proportions in APR1 are relatively higher. The row total and column average are also shown.

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 earlier in Section 2.4 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 2.6.2, this figure is 90.5%). For 73.9% of the subjects, the within-subjects correlation of contributions with guesses 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 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%).

2.6.3 Are order effects important?

Consider the order effects which distinguish Experiments 1 and 2. Table 2.3 shows the p-values in a two-sided Mann-Whitney U test. 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.

Table 2.3: P-values in Mann-Whitney U Tests.

	APR1	APR2	PUB
FOB	0.011	0.268	0.156
Contribution	0.708	0.750	0.230

2.7 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 2.5). 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 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 2.5.1).[24] 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

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

the three treatments (6 sessions per treatment).[25] 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).

Table 2.4: The frequencies of contributions in the between-subjects design.

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

Note: Figures in parentheses give the percentage of the subjects making the associated contributions.

The frequency distribution of contributions, $0, 1, ..., 10$, for the full between-subjects dataset is shown in Table 2.4. 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% level except for the APR treatment where it is significant at the 10% level.[26] However, the regression analysis in Section 2.7.4 shows that second order beliefs are a significant determinant of contributions, hence, guilt-aversion is important after all.

2.7.1 Private treatment (PR)

Of the 6 standard private sessions, four had 18 subjects each, one had 14 subjects, and one had 16 subjects[27]. 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

[25]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.

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

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

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.

2.7.2 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.

2.7.3 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.

2.7.4 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 2.5.[28]

Table 2.5: Determinants of public good contributions.

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

Note: 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).

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

[28]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.

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 2.5). 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 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 was 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.

2.8 Conclusions

Our aim in this paper is to make theoretical and empirical contributions to the literature on psychological game theory using several alternative methods of belief elicitation. We emphasize two different but possibly related emotions (Battigalli and

Dufwenberg, 2007): (1) Simple guilt-aversion/surprise-seeking, and (2) 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 where strategic interaction plays a central role.

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 the vast majority of our subjects (95%) are guilt-averse and only 5% are surprise seeking; we also find guilt-aversion at the aggregate level. In contrast, Khalmetski et al. (2015) find no aggregate guilt aversion because individual-level guilt-aversion and surprise-seeking 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.

Chapter 3

The Determinants of Gift Exchange: Theory and Evidence.

Sanjit Dhami, Mengxing Wei, Ali al-Nowaihi

Abstract

We consider two main competing explanations for the choice of effort by workers in a gift exchange game. The classical *gift exchange motive* creates an obligation on the part of the receiver of a gift to reciprocate; we consider Akerlof's formulation as well as Malmendier and Schmidt (2017) formulation. The second is based on psychological game theory (PGT), which highlights the importance of beliefs that underpin emotions such as *guilt*, and motives such as *belief-based reciprocity* in sequential games. We derive theoretical predictions of the competing models and test them against the data. Model selection tests choose Akerlof's formulation as the best in terms of parsimony and fit. However, the PGT correctly predicts the role of beliefs which the other models cannot.

3.1 Introduction

The *gift exchange game* has been fundamental in establishing the importance of *other-regarding preferences* in economics. It has led us to significantly revise our beliefs about self-regarding preferences and has led to important advances in contract theory, labour economics, and macroeconomic models among other areas (Dhami, 2016, Part 2). The main insight behind the gift exchange game in economics goes back to the seminal paper by Akerlof (1982) in the context of an employment relation between a firm and a worker, although the insight also generalizes to other contexts.

The basic idea behind the gift exchange game is that if workers are offered a rent by the firm in excess of their outside option, then they respond by putting in greater effort, independent of reputation concerns, issues of information asymmetries, and market imperfection. This insight fundamentally challenges models that are based on self-regarding workers who exhibit no reciprocity. Such workers would shirk if offered a binding wage by the firm–this prediction is refuted by the evidence (Camerer, 2003; Dhami, 2016). One may wonder if the gift-exchange behavior of the workers is simply the predicted response of self-regarding workers in the presence of efficiency wages and unemployment (Shapiro and Stiglitz, 1984). In a set of beautifully controlled experiments, Fehr et al. (1993, 1998), Fehr et al. (1997) established that controlling for the effect of confounding factors, reciprocity to a gift was observed. This has led to an explosion of interest in the literature and there are now literally hundreds of published experiments and field studies on variants of the gift exchange game that provide strong support to the idea of gift exchange.[29]

3.1.1 The competing explanations of gift exchange

In this paper we are interested in two main explanations of gift exchange.[30]

3.1.1.1 Sociological and anthropological explanations

The first explanation draws on the innate desire of humans to respond positively to a gift of another, irrespective of any other considerations; we may term this

[29]For a survey of this literature, see Dhami (2016); the basic evidence and ideas are contained in Section 5.3 of the book and several extensions and related ideas are considered generally in the introductory chapter and in Part 2 of the book. Camerer (2003) also covers a fair bit of the literature.

[30]One might also explain gift exchange using models of other-regarding preferences such as models of *inequity aversion* (Fehr and Schmidt, 1999) and models of *type-based reciprocity* (Levine, 1998). These issues have already been much studied. However, Malmendier and Schmidt (2017) show that in their setup with third party advice, the gift exchange model explains their data better than either the inequity aversion model or type-based reciprocity.

as *action-based reciprocity*. Malmendier and Schmidt (2017) very nicely capture the anthropological and sociological insights behind gift exchange: "Our evidence suggests that a gift triggers an obligation to repay, independently of the intentions of the gift giver and the distributional consequences. It seems to create a bond between gift giver and recipient, in line with a large anthropological and sociological literature on gifts creating an obligation to reciprocate."

Among the sociological and anthropological explanations, we distinguish between two kinds. The basic idea is identical in the two but the precise formulation of gift exchange differs slightly. (i) The *action-based formulation* in Akerlof (1982) that introduced this subject matter into economics, and (2) the *MS formulation*, as illustrated in the work of Malmendier and Schmidt (2017).[31]

In Akerlof's formulation, workers who receive a gift, reciprocate by working harder in proportion to the size of the gift. In the MS formulation, if the workers receive a gift, then they internalize the objective function of the gift giver into their objective function. While the differences between the two formulations are subtle, the predictions of the two models differ. In the MS formulation, an increase in wage increases the material payoff of the gift receiver (worker) but reduces the material payoff of the gift giver (firm). So the net effect of an increase in the wage on the actions of the gift receiver will depend on the values of model parameters. This is not an issue in Akerlof's formulation, where a higher wage elicits higher effort.

3.1.1.2 A psychological game theory (PGT) explanation

One possible explanation for reciprocal behavior may be that players feel *guilty* by letting down the expectations of other players. For instance, suppose that in a gift exchange game, the worker believes that the firm desires the worker to put in a certain level of effort. It might also be the case that firms who offer higher wages also expect workers to work harder. Could it be the case that the worker feels *guilty* if he puts in a lower effort relative to the one that he believes the firms wishes him to put in? Furthermore, players may try to gauge the intentions of the other player and they may engage in *belief-based reciprocity* (as distinct from action-based reciprocity, which characterizes the sociological and anthropological explanations).

The appropriate machinery to discuss guilt and belief-based reciprocity in a rigorous manner is provided by *psychological game theory* (PGT). In PGT, unlike classical game theory, beliefs directly enter into the utility functions of players and these beliefs are endogenous. A large literature supports the results that players derive utility from reciprocity and positive surprise but disutility from guilt (Dhami,

[31] We note that Englmaier and Leider (2012) propose a very similar formulation.

2016, Section 13.5). Considerations of belief-based reciprocity in PGT were introduced by Rabin (1993), drawing on the seminal contribution of Geanakoplos et al. (1989). This was extended to extensive form games by Dufwenberg and Kirchsteiger (2004)[32] and to an even more general framework by Batigalli and Dufwenberg (2009). Considerations of guilt were introduced by Battigalli and Dufwenberg (2007) and we shall be mainly interested in what they term as *simple guilt*.[33]

Example 3.1. *(Guilt and reciprocity)*[34]*: Suppose that it is common knowledge in some city that one must tip a percentage of the fare. John takes a taxi to a secluded location in the city. 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's expectations down so he pays the tip (simple guilt–aversion). Also, John gives a higher tip if he believes (John's first order belief) that the taxi driver has been pleasant to him (belief-based reciprocity).*

From the above example, guilt–aversion and reciprocity cuts in opposite direction as regards how likely John is to give a lot when the driver's expectation of a tip goes up. He is more likely under guilt–aversion, yet less likely under reciprocity in the sense that the driver then appears less kind.

3.1.2 Aims and description of the paper

The first aim of our paper is to develop the theoretical predictions of the two main classes of models described in Section 3.1.1. Application of the sociological/anthropological models to gift exchange are well known, however, there is novelty in our development of the theoretical model and its predictions particularly for the case of the MS formulation of gift exchange. Dufwenberg et al. (2011) apply PGT to public goods games but model belief-based reciprocity and simple guilt separately. Our novelty in respect of the second class of models is to apply PGT to gift exchange games and simultaneously model belief-based reciprocity and simple guilt aversion. We model belief-based reciprocity in its sequential version in Dufwenberg and Kirchsteiger (2004) and guilt as simple guilt in the form suggested in Battigalli and Dufwenberg (2007).

[32]The theoretical work of Dufwenberg and Kirchsteiger (2000) explored reciprocity and gift exchange, which considered competitions between the two workers. By contrast, our framework does not contain any interaction between the two workers.

[33]Another concept of guilt aversion, *guilt from blame*, requires the formation of third and fourth order beliefs (Battigalli and Dufwenberg, 2007). For applications of this concept, see Khalmetski et al. (2015) and Dhami et al. (2016).

[34]This is a slightly modified version of the example in Dhami (2016, Example 13.8, p. 925).

The second aim is to test the predictions of these models using new experimental data and to examine model selection issues. We need fresh experiments because we need data not just on the wage and effort levels, but also on the first and second order beliefs of the players that is critical in testing the predictions of the PGT models. *Second order beliefs* are beliefs of a player about the *first order beliefs* of other players. In principle, second order beliefs can be notoriously difficult to elicit. The reason is that if players are asked to simply state their second order beliefs, they might just ascribe to other players their own beliefs; this is the *false consensus effect.* This was empirically demonstrated by Ellingsen et al. (2010). They also suggested a solution to the problem, the *induced beliefs method*, that allows for arguably truthful revelation of second order beliefs.[35] We apply the induced beliefs method to the gift exchange game, which we believe is the first such application to this class of games.

Malmendier and Schmidt (2017) show, correctly, that if one imposes the full set of conditions of a psychological Nash equilibrium (essentially a sequential Nash equilibrium with endogenous beliefs), then we may have multiple equilibria. They show that this allows for the presence as well as the absence of gift exchange. A psychological Nash equilibrium requires the assumption of rational expectations in beliefs. There is now much evidence that in experimental settings, particularly in games played only a few times, the typical finding is one of disequilibrium beliefs (Dhami, 2016, Chapters 12, 13). For this reason, several recent papers using PGT do not impose rational expectations in beliefs (Khalmetski et al., 2015; Dhami et al. 2016). We follow these insights and do not impose equilibrium in beliefs either. This gives rise to a unique equilibrium with eminently testable restrictions. Furthermore, we are mainly interested in the effort response of the worker, conditional on the wage and the belief hierarchy in the game, rather than in the firm's choice of an optimizing wage.

The third aim, which is the least ambitious of the three, is to test if historical information can provide a focal point or an anchor for the formation of beliefs and for actions within the game. This gives some context to our experiments by introducing something akin to norms and enables us to get some perspective on how people form beliefs.

3.1.3 Main findings and organization of the paper

We find empirical support for Akerlof's action-based gift exchange formulation and the PGT model. By contrast, the MS gift exchange formulation performs relatively

[35]For other papers using the induced beliefs methodology, see Khalmetski et al. (2015) for application to dictator games and Dhami et al. (2016) application to public goods games.

poorly. In general, effort is increasing in the wage, which supports Akerlof's gift exchange formulation. In general the theoretical prediction of the MS formulation and of the PGT model on the effect of wage on effort depends on the values of the model parameters. We believe that this was not widely recognized before. Beliefs too turn out to be important. Second order beliefs are correlated with effort suggesting that simple guilt-aversion plays a role. This is not predicted by any of the gift exchange models (in our case, the Akerlof model and the MS formulation).

In order to judge the trade-off between the greater explanatory power of the PGT model against the larger number of parameters in this model, we run model selection tests. The Akaike information criterion and the Bayesian information criterion both suggest that Akerlof's action-based gift exchange formulation does best among the three models in terms of parsimony and explanatory power. Yet purely statistical criteria are just one element in model choice. Other criteria, statistical and economic, may be equally relevant. For instance, a researcher might be particularly interested in examining the role of beliefs in determining actions because he/she could be thinking of particular policy interventions to influence beliefs.

Unlike the explanation based on PGT, intentions are not important in the gift exchange explanation. Hence, players need not form beliefs about the unobserved intentions of the other players. This leads to a cognitively much simpler process in which players appear simply hard wired to reciprocate a gift or receptive to a social norm that requires them to do so. Perhaps it is this factor that gives Akerlof's gift exchange model its edge over the others.

Section 3.2 describes the basic set-up of our model. It also describes the preferences under our three different models. Sections 2.3 and 3.4 derive, respectively, the theoretical predictions of the PGT and the gift exchange models. Section 3.5 explains our experimental design. Section 3.6 pits the experimental findings against the theoretical predictions for all three models. Section 3.6.5 runs a horse race between the models by using model selection tests. We conclude in Section 3.7. Appendix B.1 contains the proofs of all the results in the papers. Appendix B.2 contains a brief note on the firm's optimal choice of wage. Appendix B.3 describes the experimental instructions, the control questions and the post–experiment survey.

3.2 The Model

Consider the gift exchange game. There is one firm (F) and two workers, $i = 1, 2$, who work on independent but identical projects and exert respective effort levels e_1 and e_2. Neither worker observes any relevant economic variable pertaining to the other worker, so social preferences among workers play no role. There are no production spillovers between the two projects, hence, the interaction between the firm and any worker is identical to a model with one firm and one worker.[36] For this reason, henceforth, we drop the subscript i for the worker and consider interaction between a single firm and a worker.

Labor is the sole factor of production and the production function of the firm, $Q : \mathbb{R}_+ \to \mathbb{R}_+$ is linear, $Q(e) = e$. Each unit of output can be sold by the firm at an exogenously given product price $p \in [0, \overline{p}]$.

3.2.1 The sequence of moves

Stage 1: The firm chooses the contractible wage level w for the worker, and $w \in [0, \overline{w}], \overline{w} \leq p$. There is also common knowledge of a norm of worker's effort level $e_N \in [0, \overline{e}]$ in similar past interactions.[37]

Stage 2: The worker observes the wage level, w, and chooses the effort level $e \in [0, \overline{e}]$. The cost function of effort of the worker, $c : \mathbb{R}_+ \to \mathbb{R}_+$ is twice continuously differentiable and is increasing and convex, $c'(e) > 0, c''(e) > 0$. In order to get closed form solutions, we shall mainly use the cost function $c(e) = \frac{e^2}{2}$; a more general convex function adds no further economic insights.

3.2.2 Beliefs in the game

Beliefs of the firm and of the worker play an essential role in psychological game theory. We now specify the beliefs in both stages.

Stage 1 beliefs: The *first order beliefs* of the firm about the Stage 2 effort of the worker are denoted by $b_F^1 \in [0, \overline{e}]$. Analogously, the first order belief of the worker about the Stage 2 wage paid by the firm is denoted by $b_W^1 \in [0, \overline{w}]$.[38] This captures the ex-ante expectation of the worker that may act like a reference point in Stage 2. In Stage 1, before the wage has been revealed to the worker, the worker intends

[36] We have introduced two workers per firm in order to increase the data for the choices made by workers, whose behavior we are mainly interested in. Thus, twice as many subjects are assigned to be workers as compared to firms.

[37] We implement this in our experiment by a public announcement that in the past, workers expended an average effort level e_N. The choice of e_N in our experiments is consistent with past findings from similar games; this is explained in more detail below when we discuss the experimental design. We did not incorporate the norm of wage, because in this way we can make the experiment as simple as possible and we are mainly interested in the behaviors of the workers.

[38] b_W^1 might be some notion of a fair wage or the norm for a wage that is consistent with the *mental model* held by the worker; such mental models may be held widely among the workers or could be peculiar to a particular worker. On mental models, see Dhami (2016, Section 19.3).

to put in a level of effort $e_0 \in [0, \overline{e}]$ in Stage 2.[39]

Stage 2 beliefs: Worker i's *second order belief* $b_W^2 \in [0, \overline{e}]$ is the belief about b_F^1. In words: this is the effort level that the worker believes that the firm expects the worker to undertake. A proper specification and measurement of second order beliefs is required to formally model and test the phenomenon of *guilt*.

Denote by B, the set of all first and second order beliefs of the firm and the worker in the two stage game, so $B = \{b_F^1, b_W^1, b_W^2\}$.[40]

It is, of course, possible that there is a distribution of beliefs and the stated first order beliefs of the firm in Stage 1 capture some summary statistic of the distribution such as the median (as in Khalmetski et al., 2015 and Dhami et al., 2016). However, in this paper we treat beliefs as point beliefs.[41]

3.2.3 Profit function of the firm

Given the assumptions made above, the profits of the firm are given by

$$\Pi = (p - w)e. \tag{3.1}$$

3.2.4 Preferences of the worker in psychological game theory (PGT)

The *material payoff* of the worker, conditional on observing the wage, w, is given by

$$\pi(w, e) = w - c(e). \tag{3.2}$$

Given the assumptions, $\pi : [0, \overline{w}] \times [0, \overline{e}] \to \mathbb{R}$ is a bounded, and twice continuously differentiable function of e.

The psychological utility of the worker $U : [0, \overline{e}] \times B \to \mathbb{R}$, includes material utility but also psychological traits such as reciprocity and guilt.[42] Its general form in PGT is given by

[39]In our experiments, we elicit e_0. The norm of worker's effort level e_N that is announced in Stage 1 may serve as an anchor for the Stage 1 beliefs of the firm (how much effort will the worker put in Stage 2, b_F^1) and for the Stage 1 beliefs of the worker (how hard the worker intends to work in Stage 2, e_0). We test empirically if this is the case later on in the paper.

[40]Since we are mainly interested in the worker's optimization problem and we are interested only in the worker's reciprocity and simple guilt, we do not need to invoke second order beliefs of the firm.

[41]Since we did not elicit subjects' distributions of beliefs in the experiments (see the *within-subjects* design in Khalmetski et al., 2015 and Dhami et al., 2016), we use point beliefs in the models.

[42]We have suppressed the role of w as an argument in U because the worker's choice is conditional on a given value of w and it is not under the worker's control.

$$U(e,B) = \pi(w,e) + Y_R\,(\theta)\,R(e,B) - Y_G\,(\theta)\,G(e,B), \tag{3.3}$$

where $Y_R \geq 0$ and $Y_G \geq 0$ are parameters that measure, respectively, the *belief-based reciprocity sensitivity* and the *guilt sensitivity* of the worker.[43] R is the worker's *belief-based reciprocity* towards the firm, G is the worker's *guilt function*, and θ captures *history–dependent framing effects* unrelated to the current economic interaction between the firm and the worker. Unless we need to discuss framing issues, we will simply suppress the dependence of Y_R and Y_G on θ. As we shall see below, PGT highlights *belief-based reciprocity*, while the classical definition of gift exchange highlights the role of *action-based reciprocity* based on a norm of gift exchange. This separates the sense in which we use the term 'reciprocity' in these two different models.

The guilt function of the worker is given by the following bounded function

$$G(e,B) = max\left\{0, b_W^2 - e\right\}. \tag{3.4}$$

In (3.4), guilt takes the form of *simple guilt*, and it captures the extent to which a player derives disutility from falling behind the expectations of the other player (Battigalli and Dufwenberg, 2007). For pedagogical ease, this function is asymmetric, so there is no extra utility from exceeding the expectations of the other player.[44]

In our experiments, we introduce *history–dependent framing effects* through different values of $\theta = U, N, K$. We introduce three possible history frames. In the *neutral history frame* ($\theta = N$), no historical information is given about the firm. In the *kind history frame*, $\theta = K$ (respectively, the *unkind history frame*, $\theta = U$), workers are told that the firm, in the past, had clearly exhibited kind (respectively

[43] The objective function in (3.3) applies to a single individual who is characterized by a pair of parameters (Y_R, Y_G). In the population, and in our experiments, there is likely to be a distribution of Y_R, Y_G values drawn from some joint distribution that is defined over $\mathbb{R}_+ \times \mathbb{R}_+$. We know of no way of directly empirically observing Y_R, Y_G for an individual but we make predictions conditional on a fixed pair (Y_R, Y_G) and also state our comparative static results in terms of the conditions on the magnitudes of these parameters.

[44] Indeed, players may derive utility from positively surprising the other players. This *surprise-seeking motive*, the flip side of the guilt aversion motive, is theoretically modelled and empirically tested in Khalmetski et al. (2015) and Dhami et al. (2016). Since there may be a very small proportion of people who are relatively surprise-seeking, we do not incorporate it in this paper.

unkind) intentions towards its workers.[45] It is plausible that framing effects in (3.3) may be captured through the following inequalities.

$$\begin{cases} Y_R(U) < Y_R(N) < Y_R(K) \\ Y_G(U) < Y_G(N) < Y_G(K) \end{cases}. \tag{3.5}$$

In this version, the impact of framing is solely on the reciprocity sensitivity and the guilt sensitivity parameters of the worker. In the kind history treatment, the firm is revealed to be kind ($\theta = K$) in its previous interactions with workers. This may elicit even greater belief-based reciprocity on the part of the worker and even greater guilt from letting down the expectations of the firm relative to the neutral history treatment, hence, $Y_R(N) < Y_R(K)$ and $Y_G(N) < Y_G(K)$. The converse is true when the firm is revealed to be unkind ($\theta = U$) in the past. The neutral history treatment serves as a benchmark with which we can compare our results.

We now apply the framework of Rabin (1993) (for simultaneous move games) and Dufwenberg and Kirchsteiger (2004) (for sequential games) to specify the belief-based reciprocity term $R(e, B)$ in (3.3). Our game has a sequential structure, so we mainly follow the Dufwenberg-Kirchsteiger method. In this framework,

$$R(e, B) = k_{WF}\widehat{k}_{FW}, \tag{3.6}$$

where k_{WF} is the *kindness of the worker to the firm, as perceived by the worker* and $\widehat{k}_{FW}$ is the *kindness of the firm to the worker as perceived by the worker.* Let us now explain these two functions.[46] This is the sense in which reciprocity is belief-based. If firm is perceived to be kind ($\widehat{k}_{FW} > 0$), then by reciprocating the kindness ($k_{WF} > 0$), the worker increases utility as given in (3.3). Similarly, utility can be increased by reciprocating unkindness ($\widehat{k}_{FW} < 0$) with unkindness ($k_{WF} < 0$).

Let us first compute k_{FW}. This requires the specification of an *equitable payoff* to the worker, $\overline{\pi}$:

$$\overline{\pi}\left(w, b_F^1\right) = \frac{1}{2}\max\left\{\pi(w, b_F^1),\, w \in [0, \overline{w}]\right\} + \frac{1}{2}\min\left\{\pi(w, b_F^1),\, w \in [0, \overline{w}]\right\}. \tag{3.7}$$

At the time of announcement of the wage, the firm does not know the effort level

[45] Details are given in the description of the experiments below, however, the kindness/unkindness of the firm is revealed in a context unrelated to a gift-exchange situation. There is evidence that the revelation of behavior/intentions in one domain may influence inferences about behavior in another domain. For instance, in trust games, Charness et al. (2011) found that their subjects use past data on trust to infer greater trustworthiness. Whether this turns out to be the case in the gift exchange game is an empirical question that we try to give the answer to.

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

chosen by the worker, so it uses its first order beliefs about the effort, b_F^1. The equitable payoff is an equally weighted average of the maximum and the minimum payoffs that the firm can guarantee the worker through its choice of a wage rate. Since the material payoff of the worker is linear in w, the worker's highest possible material payoff arises when $w = \overline{w}$ and the lowest possible payoff arises when $w = 0$. Thus, using (3.2) in (3.7), we get that[47]

$$\overline{\pi}\left(w, b_F^1\right) = \frac{1}{2}\overline{w} - c\left(b_F^1\right). \tag{3.8}$$

Then, we define k_{FW} as follows.

$$k_{FW}\left(w\right) = \pi(w, b_F^1) - \overline{\pi}\left(w, b_F^1\right), \tag{3.9}$$

i.e., the firm is kind to the worker if through its choice of a wage level, w, it gives the worker a material payoff greater than the equitable payoff. Otherwise the firm is unkind. Substituting (3.2), (3.7) in (3.9), we get

$$k_{FW}\left(w\right) = w - \frac{1}{2}\overline{w}. \tag{3.10}$$

The worker needs to form inferences about $k_{FW}\left(w\right)$ after observing the wage but before choosing the effort level in Stage 2. In general, the worker does not know k_{FW}, so he need to form inferences about it. However, given the structure of the problem, all variables on the RHS in (3.12) are in the information set of the worker in Stage 2. Hence, the worker's perception of the firm's kindness is completely accurate, so

$$\widehat{k}_{FW}\left(w\right) = w - \frac{1}{2}\overline{w}. \tag{3.11}$$

We now need to compute the kindness of the worker to the firm as perceived by the worker, k_{WF}. We first need to compute the equitable payoff of the firm, $\overline{\Pi}\left(w, e\right)$. In Stage 2, before the worker makes the effort choice, the wage offered by the firm is already observed, so by Bayes' rule the worker must place probability 1 on the observed wage (this is a part of the requirement of sequential rationality). Proceeding as in the case of (3.7), we have

$$\overline{\Pi}\left(w, e\right) = \frac{1}{2}\max\left\{\Pi\left(w, e\right), \, e \in [0, \overline{e}]\right\} + \frac{1}{2}\min\left\{\Pi\left(w, e\right), \, e \in [0, \overline{e}]\right\}. \tag{3.12}$$

[47] In (3.8), the first order beliefs b_F^1 do not depend on the wage. The reason is that in each subgame it is still the case that the worker gets the maximum payoff if $w = \overline{w}$ and the minimum if $w = 0$. Furthermore, k_{FW} in (3.10) is independent of the effort level.

Π, defined in (3.1), is linear in e. Hence, for any w, the firm's profit function Π takes its maximum value when $e = \overline{e}$ and its minimum when $e = 0$. It follows that we can rewrite (3.12) as follows.

$$\overline{\Pi}(w, e) = \frac{1}{2}(p - w)\overline{e}. \tag{3.13}$$

The kindness of the worker towards the firm, k_{WF}, is given by

$$k_{WF}(e) = \Pi(w, e) - \overline{\Pi}(w, e), \tag{3.14}$$

and it depends on the effort level chosen by the worker. If the effort choice of the worker ensures that the profits of the firm are greater than the equitable payoff of the firm then the worker is kind to the firm. Otherwise the worker is unkind to the firm. Substitute (3.1) and (3.13) in (3.14), we get

$$k_{WF}(e) = (p - w)\left(e - \frac{1}{2}\overline{e}\right). \tag{3.15}$$

Substituting (3.11) and (3.15) in (3.6) we get

$$R(e, B) = (p - w)\left(w - \frac{1}{2}\overline{w}\right)\left(e - \frac{1}{2}\overline{e}\right) \tag{3.16}$$

The interpretation of the belief-based reciprocity term in (3.16) is very intuitive. By assumption, $w \leq p$, so the sign of $R(e, B)$ is determined by the product $\left(w - \frac{1}{2}\overline{w}\right)\left(e - \frac{1}{2}\overline{e}\right)$. If the worker gets a wage higher than $\frac{1}{2}\overline{w}$, which is interpreted as a kind offer, then he responds by putting in an effort level higher than $\frac{1}{2}\overline{e}$ to reciprocate the firm's kindness. Analogously, unkind offers ($w < \frac{1}{2}\overline{w}$) by the firm are reciprocated by lower effort choices ($e < \frac{1}{2}\overline{e}$).

Substituting (3.16) and (3.4) in (3.3), we find that the objective of the worker is to choose the effort level, e, conditional on the wage, w, and the frame, θ, in order to

$$\underset{\langle e \in [0,\overline{e}] | w, \theta \rangle}{Max} \; U = \begin{cases} w - c(e) + Y_R(\theta)(p - w)\left(w - \frac{1}{2}\overline{w}\right)\left(e - \frac{1}{2}\overline{e}\right) - Y_G(\theta)(b_W^2 - e) & if \quad e < b_W^2 \\ w - c(e) + Y_R(\theta)(p - w)\left(w - \frac{1}{2}\overline{w}\right)\left(e - \frac{1}{2}\overline{e}\right) & if \quad e \geq b_W^2 \end{cases}, \tag{3.17}$$

Given the assumptions, U is a strictly concave function. It is twice continuously differentiable except at the point $e = b_W^2$. Hence, it reaches its unique maximum in the interval $[0, \overline{e}]$. We shall only be interested in the properties of the effort response of the worker. Issues about the optimal choice of a wage level by the firm are briefly commented on in Appendix B.2.

We now comment on three important features of our problem:

1. A large body of research suggests that the beliefs of the players are not in equilibrium, at least for games that are not played a very large number of time.[48] Indeed, requiring sequential rationality in psychological games may lead to multiple equilibria.[49] For this reason, we do not require sequential rationality of actions with beliefs as in Khalmetski et al. (2015) and Dhami et al. (2016).

2. Note the absence of a participation constraint for the worker in our formulation, which might appear to be at odds with the formulation of a typical agency problem. The reason for this is that our data comes from an experimental setting in which workers make a voluntary participation decision, having been promised a participation fee that reflects their outside option. If effort is onerous, then conditional on the wage, workers can implement a low or zero effort level, take their participation fee and walk away. Hence, an explicit participation constraint is not needed. This comment also applies to the preferences under gift exchange that we consider next. The experimental instructions were common to all players, so firms observed the objective function of workers and vice-versa. Thus, there are no issues of asymmetric information and an incentive compatibility constraint is not needed. This remark also applies to our formulation of the problem under the gift exchange theories.

3. Dufwenberg et al. (2011) consider the two cases of belief-based reciprocity only ($Y_R > 0$ but $Y_G = 0$) and simple guilt only ($Y_R = 0$ but $Y_G > 0$). Our general model introduces both; see (3.17). In our results below, the two special cases can simply be recovered by setting one of Y_R and Y_G equal to zero.

3.2.5 Preferences under gift exchange– I (Akerlof's formulation)

Let us now consider the preferences of the worker in the classical gift-exchange case. There is no economic or social exchange between the two workers in our experiment so no gift exchange can take place between them. The only relevant gift

[48]In the context of psychological games, see Khalmetski et al. (2015), Dhami et al. (2016) and for the general context, see Camerer (2003) and Dhami (2016, Parts 5,6). Even in games that are played a very large number of times, unless the underlying economic environment is stationary, there is no guarantee that beliefs will be in equilibrium (Camerer, 2003; Dhami, 2016, Parts 5,6).

[49]For the original contribution, see Geanakoplos et al. (1989); for an application to simultaneous move games, see Rabin (2003); and for an application to gift exchange see Malmendier and Schmidt (2017).

exchange takes place between a firm and its workers. In the original description of gift exchange Akerlof (1982) writes (p. 544): "On the worker's side, the "gift" given is work in excess of the minimum work standard; and on the firm's side the "gift" given is wages in excess of what these women could receive if they left their current jobs." In particular the role of intentionality of actions captured by considering the beliefs of the players that play a critical role in defining belief-based reciprocity (see, for instance, (3.9)) do not play a role here. Akerlof's formulation of the gift exchange motive suggests that the worker maximizes the following utility function.

$$\underset{\langle e\in[0,\overline{e}]\rangle}{Max}\, W = w - c(e) + \gamma\left(\theta\right)\left(w - w_0\right)\left(e - e_{\min}\right),\ \gamma\left(\theta\right) > 0. \tag{3.18}$$

In (3.18), the third term captures gift exchange. It is weighted by $\gamma\left(\theta\right)$, which has as its argument, the treatment information $\theta = U, N, K$. w_0 is the outside option of the worker and $e_{\min}$ is the minimum effort level. Let us normalize $w_0 = e_{\min} = 0$; this does not alter the economic insights but just relocates the origin. Thus, we can rewrite (3.18) as

$$\underset{\langle e\in[0,\overline{e}]\rangle}{Max}\, W = w - c(e) + \gamma\left(\theta\right) we,\ \gamma\left(\theta\right) > 0. \tag{3.19}$$

When the firm pays the worker a higher wage, it increases the marginal utility of effort. Importantly, this effect is independent of intentions or distributional consequences. Of all the competing models that we consider, this is cognitively the simplest formulation.

We may capture the framing effects caused by providing the historical information in the same manner as in (3.5) by postulating the following reasonable hypothesis.

$$\gamma\left(U\right) < \gamma\left(N\right) < \gamma\left(K\right). \tag{3.20}$$

Thus, gift exchange is likely to be more salient in the Kind history treatment and least salient in the Unkind history treatment. When it is not important to highlight the treatment effect, we may write $\gamma\left(\theta\right)$ as simply γ.

3.2.6 Preferences under gift exchange– II (MS formulation)

Malmendier and Schmidt (2017) give a new formulation of gift exchange. We specify the utility of the worker as follows:

$$V = \pi(w, e) + \gamma\left(\theta\right)\left(w - b_W^1\right)\Pi, \tag{3.21}$$

where $\gamma(\theta)$ plays the same role that it plays in (3.18) and (3.20). Our formulation is consistent with Assumption 1 in Malmendier and Schmidt (2017). They describe their analogue of b_W^1 as the (possibly mixed) strategy profile that players expect to be played in the game under consideration, e.g., because of past experience in similar circumstances, or because it constitutes a social norm, or because it is an equilibrium of the game that players expect to be played. Indeed, our experimental results show that the distribution of b_W^1 and w is statistically indistinguishable, suggesting that the use of b_W^1 for some mutually held norm may have merit.

From (3.21), the nature of gift exchange between the firm and the worker is described as follows. If $w - b_W^1 > 0$, i.e., the worker gets a higher than expected wage, then worker $i = 1, 2$ places a positive weight on the profits of the firm that arise from hiring the worker, Π. Conversely, if $w - b_W^1 < 0$, then a negative weight is accorded to Π. Substitute π and Π in (3.21) to get

$$\underset{\langle e\in[0,\overline{e}]\rangle}{Max}\, V = w - c(e) + \gamma(\theta)\,(w - b_W^1)(p - w)e \tag{3.22}$$

Englmaier and Leider (2012) also have a similar formulation, although they do not consider the role of beliefs in the sense of our paper.

3.3 Effort choice under belief-based reciprocity and simple guilt

Differentiating U with respect to e in (3.17), conditional on the firm's choice of wage, w, we get

$$\frac{\partial U}{\partial e} = \begin{cases} -e + Y_R(\theta)\,(p-w)\left(w - \frac{1}{2}\overline{w}\right) + Y_G & if \quad e < b_W^2 \\ -e + Y_R(\theta)\,(p-w)\left(w - \frac{1}{2}\overline{w}\right) & if \quad e \geq b_W^2 \end{cases}. \tag{3.23}$$

In any of the rows on the RHS of (3.23), the first term is the marginal cost from an additional unit of effort. The second term in any row is the marginal effect of belief-based reciprocity. A higher reciprocal sensitivity, Y_R, or a greater increment in wage over $\frac{1}{2}\overline{w}$ increases marginal utility. The magnitude of guilt sensitivity, as captured by Y_G, enhances worker's marginal utility from extra effort when it is below the expectation of the other player (last term in the first row of (3.23)). Since $\frac{\partial^2 U}{\partial e^2} = -c''(e) < 0$ for any $e \in [0, \overline{e}]$, the utility function is strictly concave.

Since $w, \overline{w}$ are fixed at the time the worker makes the effort choice, we consider separately the domains of positive belief-based reciprocity ($w > \frac{1}{2}\overline{w}$) and negative belief-based reciprocity ($w < \frac{1}{2}\overline{w}$). We denote the optimal effort choice under belief-based reciprocity and simple guilt by e^*.

3.3.1 Negative belief-based reciprocity ($w - \frac{1}{2}\overline{w} < 0$)

In the domain of negative belief-based reciprocity, $w - \frac{1}{2}\overline{w} < 0$, i.e., the worker receives less than the equitable wage so the firm's intentions are perceived by the worker to be unkind. From (3.23), when $e \geq b_W^2$, we have $\frac{\partial U}{\partial e} < 0$, so it is optimal to choose the smallest effort level in this interval, $e = b_W^2$. However, whenever $e < b_W^2$ (first row of (3.23)), there is an additional tradeoff. The worker would like to conditionally reciprocate a lower than equitable wage offer by putting in a low effort. However, guilt from falling behind the expectations of the firm pushes the worker in the direction of greater effort. The outcome is a tug-of-war between belief-based reciprocity and simple guilt–aversion. If the guilt–aversion parameter, Y_G is relatively small (respectively, relatively large) then optimal effort is zero (respectively, as large as possible). For intermediate levels of guilt, we get an interior solution in the interval $[0, b_W^2]$. These insights are formalized in the next proposition.

Proposition 3.1. *: Consider the case of negative belief-based reciprocity, $w - \frac{1}{2}\overline{w} <$ 0.*
(i) If guilt aversion is low enough, in the sense that $Y_G < -Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right)$, then the optimal effort choice is $e^* = 0$.
(ii) If guilt aversion is high enough in the sense that $Y_G \geq b_W^2 - Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right)$, then the optimal effort choice is $e^* = b_W^2$.
(iii) For intermediate levels of guilt aversion,

$$-Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right) < Y_G < b_W^2 - Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right),$$

we have an interior solution to effort given by

$$0 < e^* = Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right) + Y_G < b_W^2.$$

We have $\frac{\partial e^*}{\partial w} \gtreqqless 0$ if $p + \frac{1}{2}\overline{w} - 2w \gtreqqless 0$. Effort is predicted to be highest in the Kind history treatment ($\theta = K$), lowest in the Unkind history treatment ($\theta = U$), and intermediate in the Neutral history treatment ($\theta = N$).

Corollary 1. *: If Y_R and Y_G are common across all subjects in the experiment, then there is a positive correlation between optimal effort, e^*, and second order beliefs, b_W^2.*

The following testable restrictions arise from the case of negative belief-based reciprocity ($w - \frac{1}{2}\overline{w} < 0$).

1. The optimal effort level is never higher than the second order beliefs of the worker: The highest optimal effort level is achieved in the case of Proposition 3.1(ii), which is $e^* = b_W^2$.

2. There is a positive correlation between optimal effort and second order beliefs, b_W^2 (Corollary 1).

3. The effect of the wage w on optimal effort is ambiguous: In Proposition 3.1(i), (ii) w does not influence optimal effort. However, in Proposition 3.1(iii), the effect of the wage w on optimal effort is ambiguous and depends on the parameter values.

4. First order beliefs of the worker, b_W^1, do not influence the optimal effort level. This result stands in contrast to those in Dufwenberg et al. (2011) where public goods contributions of players are influenced by the first order beliefs of the players about the contributions of other players. The reason is that the public goods game is a simultaneous move game and players do not observe the contributions of others when they choose their own contributions. In contrast, in our sequential game, workers observe the wage announced by the firm before they choose their effort level. Sequential rationality implies that they should place a probability 1 on the observed choice of the wage rate. Thus, their first order beliefs about the wage, b_W^1, are irrelevant.

3.3.2 Positive belief-based reciprocity ($w - \frac{1}{2}\overline{w} > 0$)

Next we consider the case of positive belief-based reciprocity, $w - \frac{1}{2}\overline{w} > 0$. In this case, the belief-based reciprocity motive and the simple guilt–aversion motives both push the worker in the direction of greater effort level. Indeed, if the belief-based reciprocity parameter, Y_R, is high enough then the worker might exert the maximum possible effort, $\overline{e}$. Depending on the size of the belief-based reciprocity parameter, we might have interior or corner solutions in the two intervals $[0, b_W^2)$ and $[b_W^2, \overline{e}]$. We now formalize these insights in the next proposition.

Proposition 3.2. *: Consider the domain of positive belief-based reciprocity ($w - \frac{1}{2}\overline{w} > 0$).*
(i) If $Y_R \geq \frac{\overline{e}}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$, then the optimal effort level is given by $e^* = \overline{e}$.
(ii) If $\frac{b_W^2}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)} < Y_R < \frac{\overline{e}}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$, then the optimal solution is $e^* = Y_R(p -$

$w)\left(w-\frac{1}{2}\overline{w}\right) > b_W^2$. Effort is predicted to be highest in the Kind history treatment ($\theta = K$), lowest in the Unkind history treatment ($\theta = U$), and intermediate in the Neutral history treatment ($\theta = N$).
(iii) If $\frac{b_W^2 - Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)} \leq Y_R \leq \frac{b_W^2}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$, then the optimal solution is $e^* = b_W^2$.
(iv) If $Y_R < \frac{b_W^2 - Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$, then the optimal solution is $e^* = Y_R(p-w)\left(w-\frac{1}{2}\overline{w}\right) + Y_G < b_W^2$. Effort is predicted to be highest in the Kind history treatment ($\theta = K$), lowest in the Unkind history treatment ($\theta = U$), and intermediate in the Neutral history treatment ($\theta = N$).
(v) In cases (ii) and (iv), we have $\frac{\partial e^*}{\partial w} \gtreqless 0$ if $p + \frac{1}{2}\overline{w} - 2w \gtreqless 0$.

Corollary 2. *: If Y_R and Y_G are common across all subjects in the experiment, then there is positive correlation between optimal effort, e^*, and second order beliefs, b_W^2.*

Since we cannot observe Y_R, Y_G, it is not possible to test all the predictions in Proposition 3.2. The testable restrictions in the case of positive reciprocity are as follows.

1. Recall that under negative reciprocity, e^* cannot exceed b_W^2 (Proposition 3.1(ii)). Depending on the strength of the reciprocity parameter, under positive reciprocity, e^* can exceed b_W^2 (Proposition 3.2(i), (ii)).

2. The optimal effort level is positively correlated with second order beliefs b_W^2 (Corollary 2).

3. The optimal effort level is independent of first order beliefs b_W^1 as in the case of the negative belief-based reciprocity.

4. Whenever there is an interior solution to optimal effort, excluding $e^* = b_W^2$ (i.e., $e^* \in (0, \overline{e}) / \{b_W^2\}$), then the effect of wage on optimal effort is ambiguous; see Proposition 3.2(v).

3.4 Gift exchange and optimal effort choice

In this section, we consider the two models of gift exchange that we have outlined above.

3.4.1 Gift exchange– I (Akerlof's formulation)

W given in (3.19) is a strictly concave function of e and it is continuous on a compact set, so a maximum exists and it is unique. Differentiating W with respect to e, we get $\frac{dW}{de} = -e + \gamma(\theta) w$. At $e = 0$ we have $\frac{dW}{de} > 0$. Hence the solution lies in the set $(0, \overline{e}]$. Solving out for e and denoting the optimal solution by e^G, we get

$$e^G = \min\left\{\gamma(\theta) w, \overline{e}\right\}. \tag{3.24}$$

An immediate implication is that when e^G is an interior solution, it is strictly increasing in the wage rate, w. We summarize these results next.

Proposition 3.3. *(Action-based gift exchange): Consider the worker's utility function W in (3.19) under Akerlof's formulation of the gift exchange problem. Then, we have the following results:*
(i) For positive wage, optimal effort is strictly positive and given by (3.24).
(ii) When optimal effort lies in the interval $(0, \overline{e})$, it is strictly increasing in the wage, w.
(iii) The beliefs of the players do not influence the effort level.
(iv) If $e^G \in (0, \overline{e})$ for $\theta = K, N, U$, then we have that effort is highest in the Kind history treatment ($\theta = K$), lowest in the Unkind history treatment ($\theta = U$), and intermediate in the Neutral history treatment ($\theta = N$).

The result in Proposition 3.3 follows by using (3.20) in $e^G = \gamma(\theta) w$.

3.4.2 Gift exchange– II (MS formulation)

Differentiating V, given in (3.22), with respect to e, we get:

$$\frac{\partial V}{\partial e} = -e + \gamma(\theta)(w - b_W^1)(p - w). \tag{3.25}$$

V is a continuous, strictly concave function on a compact set, so a maximum exists and it is unique. In particular, the first order condition is sufficient. From (3.25), we may distinguish between two possibilities. Negative gift exchange ($w - b_W^1 \leq 0$) and positive gift exchange ($w - b_W^1 > 0$). The main testable implications of this model are given in the next proposition. Denote the optimal effort by e^{G1}.

Proposition 3.4. *(i) In the domain of negative gift exchange, optimal effort is zero ($e^{G1} = 0$).*
(ii) In the domain of positive gift exchange, optimal effort, e^{G1}, is strictly positive. It is given by

$$e^{G1} = \min\left\{\gamma(\theta)(w - b_W^1)(p - w), \overline{e}\right\}. \tag{3.26}$$

When the solution is interior, then (a) the effect of w on the optimal effort is ambiguous, and (b) effort is decreasing in b_W^1. In particular, we have $\frac{\partial e^{G1}}{\partial w} \gtreqless 0$ if

$p + \frac{1}{2}\overline{w} - 2w \gtreqless 0$. *(c) If $e^G \in (0, \overline{e})$ for $\theta = K, N, U$, we have that effort is highest in the Kind history treatment ($\theta = K$), lowest in the Unkind history treatment ($\theta = U$), and intermediate in the Neutral history treatment ($\theta = N$).*

Proposition 3.4 gives the testable implications. When wage is lower than initially expected ($w - b_W^1 \leq 0$) there is no gift exchange and we are back to the neoclassical prediction of a zero optimal effort level. However, when wage is higher than expected ($w - b_W^1 > 0$ and $w < p$), then effort is strictly positive and an interior solution obtains. However, in this case we cannot sign the effect of w on effort. To see this, recall that the second term on the RHS of (3.22) is $\gamma(\theta)(w - b_W^1)(p - w)e$ and the marginal effect of this term in (3.25) is $\gamma(\theta)(w - b_W^1)(p - w)$, which could be positive or negative. Thus, on the one hand, wage increases the worker's utility but on the other hand it reduces the firm's profits, leaving the net effect unclear.

3.5 Experimental Design

There are three treatments in our design that reflect the role of history– Neutral, Kind and Unkind; 5 sessions were conducted under each treatment in the SelLab in China. The only difference among the treatments was the additional historical background information. In the Neutral treatment, no background information was provided. In the Kind treatment, all subjects received the following background information: "In the past, the firm in your group not only approved worker's paid leave but also extended it by two days." In the Unkind treatment, the following background information was provided to all subjects: "In the past, the firm in your group found an excuse to put off worker's paid leave."

No subjects participated in more than 1 treatment. A total of 246 undergraduate and postgraduate students participated in the experiments; no one participated in more than one session. The computer programs were written using the software z-Tree (Fischbacher, 2007).

At the beginning of the experiment, each subject randomly drew a labelled ball from an opaque box to determine their role in the experiment; balls were labelled either as firms or employees. Each firm was matched with two employees in an anonymous manner. All the subjects received a copy of the instructions, and the experimenter read the instructions aloud in front of them prior to the commencement of the experiment. This also made it likely that there was common knowledge of the game and of the payoffs.

There were two stages in each session that are designed to implement the two stages in our theoretical model described in Section 3.2.1.

Stage 1: In the first stage, each firm decided on the wage level to pay, which in total should not exceed their endowment of 200 tokens (experimental currency). Each worker's wage was restricted to multiples of 10 tokens, i.e. $\{0, 10, \cdots, 100\}$.[50] The effort of the workers was restricted to the set $\{0.1, 0.2, \cdots, 1.0\}$. Firms were required to guess the possible effort levels that could be chosen by their workers from this set. It was also announced publicly that workers *normally* expended the effort level 0.4 in past experiments; this is consistent with the findings of previous experiments.[51]. We wished to check if information on the average past effort levels could serve as a norm around which the beliefs of firms and workers may be anchored. Workers were required to state their *intended effort* and their expectation of the wage they were likely to receive in Stage 2; workers were assured of the confidentiality of this information.

We follow the *induced-beliefs methodology* of Ellingsen et al. (2010), replicated further by Khalmetski et al. (2015) and Dhami et al. (2016), to inform one side in the transaction about the beliefs of the other side. We take care, as is required in the method, to ensure that the firm does not know that its beliefs will be transmitted to the other party (in Stage 2; see below).[52] This is likely to minimize or eliminate strategic elements in affecting the beliefs of the firm. If beliefs are point beliefs and are not stated strategically, then, the firm's first order beliefs and the worker's second order beliefs may be expected to coincide. In terms of our theoretical model, this would imply that

$$b_W^2 = b_F^1.$$

To minimize the possibility of random/inaccurate guesses, we incentivized beliefs by announcing that all subjects had a chance to win an additional 5 *Yuan* based on the accuracy of their guesses. At the end of the experiment, we randomly chose one firm and one worker among the set of subjects whose guess was correct. If nobody guessed accurately, then we randomly chose one among those with a guess that was the nearest to the actual choice and gave away a prize of 2 *Yuan*.[53]

Stage 2: In the second stage, workers were informed of their wage (chosen by the firm in Stage 1) and also the firm's expectation of their effort (chosen by the firm in Stage 1). Following the induced beliefs design, the workers were informed of the

[50]Since each firm was matched with two workers, it could theoretically exhaust its endowment of 200 tokens by offering each of the two workers 100 tokens each.

[51]$\leq$Previous results using the gift exchange game show that the average effort level chosen by the subjects is around 40% of the maximal level (Fehr et al., 1993; Fehr et al., 1998; Charness et al., 2004; Charness and Kuhn, 2007; Gächter et al., 2013).

[52]Technically, no deception is involved because the firm is not told that its beliefs will or will not be passed on to the other party.

[53]This setting is incentive compatible and in line with Ellingsen et al. (2010) and Khalmetski et al. (2015).

Table 3.1: Cost function of the worker

Effort	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Cost	0	1	2	4	6	8	10	12	15	18

firm's expectation of their effort, before the workers chose their actual effort level. Besides, the workers also received the information that "your expectation of firm's wage is not revealed to the firm and the firm does not know that its expectation of effort is revealed to the worker." This setting minimizes the likelihood that the worker's third or higher order beliefs come into play.

Once the workers had received this information, they could choose the *actual effort* from the set $\{0.1, 0.2, \cdots, 1.0\}$. Workers and firms were informed that the *actual effort* level would be conveyed to the firm in the end of the experiment.

The firm's profit function from hiring any worker is given by $\pi_F = (100 - w) \times e$ and the worker's material payoff is given by $w - c(e)$. The cost function, $c(e)$, is shown in Table 3.1.[54] The profit function and the cost function were common knowledge.

Subjects could commence with the experiment only if they correctly answered four control questions that testing their understanding of the experimental design. At the end of the experiment, subjects were paid individually and privately, at the exchange rate of 2 Tokens= 1 Yuan.

3.6 Experimental Results

3.6.1 Some descriptive statistics

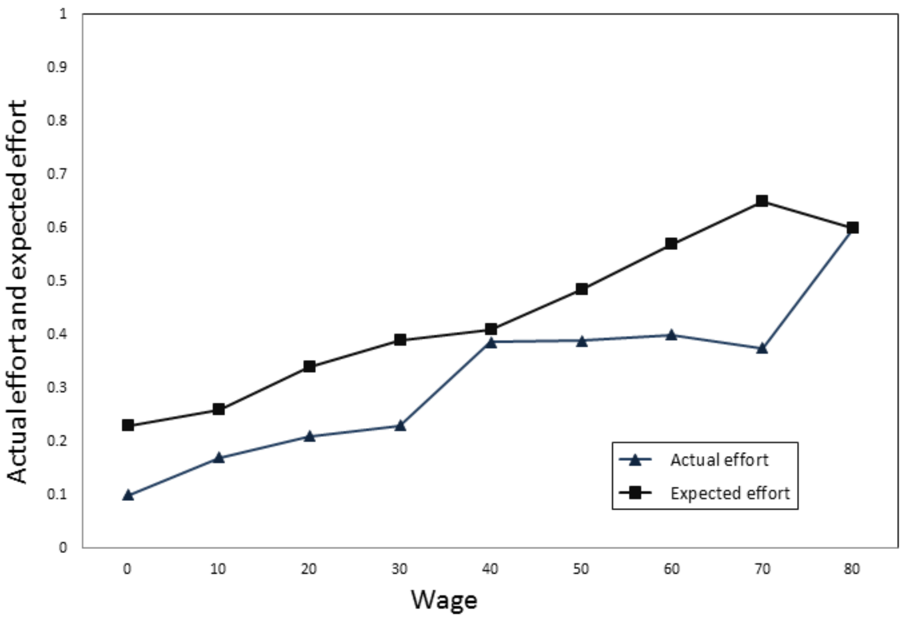
Figure 3.1 shows that actual and desired effort levels are increasing in wages; it plots the average values corresponding to each level of wage. In experimental settings, participants receive a participation-fee that reflect the opportunity cost of time. Any additional payments, such as wages in our experiment, reflect rents from the game. Thus, Figure 3.1 reproduces the gift-exchange graph that shows: (1) Actual effort is increasing in rents. (2) Actual effort is below the expected effort, so there is a degree of shirking relative to expectations of the firm.[55]

The results in this section are expressed in terms of our experimental currency 'tokens' or in percentages. We begin by giving some descriptive statistics on the

[54]The profit function and cost functions followed Fehr et al. (1993) who set the lowest effort level to 0.1 (rather than 0), and it costs zero so it is the counterpart of the zero effort level in our theoretical model.

[55]In our data, we did not have any observations for a wage level of 90 and only one observation for a wage of 100, which we have omitted.

Figure 3.1: Actual and expected effort plotted against wage.



average wage, actual effort, and first order belief across the three treatments in Table 3.2.

Table 3.2: Average actions and first order beliefs.

Treatment	Firm		Worker	
	Wage	FOB	Effort	FOB
Neutral	42.50	0.43	0.33	40.89
Unkind	42.42	0.45	0.33	40.30
Kind	40.00	0.45	0.36	40.24

Recall that the different treatments provide historical information on variables unrelated to the current experiment but those that might reveal the intentions of the firm in another context (neutral, kind and unkind). The differences among the treatments in Table 3.2 are quite small. Thus, the provision of information of this sort does not influence outcomes and beliefs. This speaks to the question of the portability of intentions across contexts.

Each of our three models predicts that, *if kindness intentions are portable across contexts*, then optimal effort will be highest in the Kind treatment, lowest in the Unkind treatment and intermediate in the Neutral treatment (Propositions 3.1, 3.2, 3.3, 3.4). We find only very limited support for this hypothesis and on balance, we must conclude that kindness intentions are not portable across treatments. The average effort levels after omitting the highest effort and the lowest effort levels, are statistically indistinguishable across the three treatments. The average effort levels are: 0.417 for Neutral (41 observations); 0.422 for Unkind (45 observations); 0.437 for

Kind (30 observations). We use a two-sided $t-$test to test if the differences in mean effort between treatments are significant: Neutral minus Unkind ($p = 0.546$); Unkind minus Kind ($p = 0.73$); Neutral minus Kind ($p = 0.34$). None of the differences is significantly different. We also use the nonparametric Mann-Whitney U test to see if the distributions of effort across the treatments are significantly different: Neutral minus Unkind ($p = 0.779$); Unkind minus Kind ($p = 0.63$); Neutral minus Kind ($p = 0.485$). None of the distributions is significantly different from the other.

For these reasons, we shall pool the data across the treatments in many of the cases that we report below.[56]

Remark 3.1. *: From Propositions 3.1, 3.2 and 3.4, we have* $\frac{\partial e}{\partial w} \gtreqless 0$ *if* $p + \frac{1}{2}\overline{w} - 2w \gtreqless 0$. *Thus, the predicted effect of wage on effort is ambiguous under psychological game theory and under the MS formulation of gift exchange. In contrast, the effect of wage on effort is unambiguously positive in the Akerlof's gift exchange model (Proposition 3.3). However, in our data we have only two cases of* $p + \frac{1}{2}\overline{w} - 2w < 0$. *In all the remaining cases we have* $p + \frac{1}{2}\overline{w} - 2w > 0$, *hence, the prediction* $\frac{\partial e}{\partial w} > 0$ *for our three theories turns out to be practically identical for our data. Thus, on this count (the effect of wages on effort) we are unable to stringently test psychological game theory and the MS formulation of gift exchange.*

We introduce information about how much effort workers have exerted in the past in similar situations–a sort of an effort norm. It is plausible that the provision of this information acts as an anchor for the formation of beliefs in the experiment.[57]

Table 3.3: Percentages of following the norm.

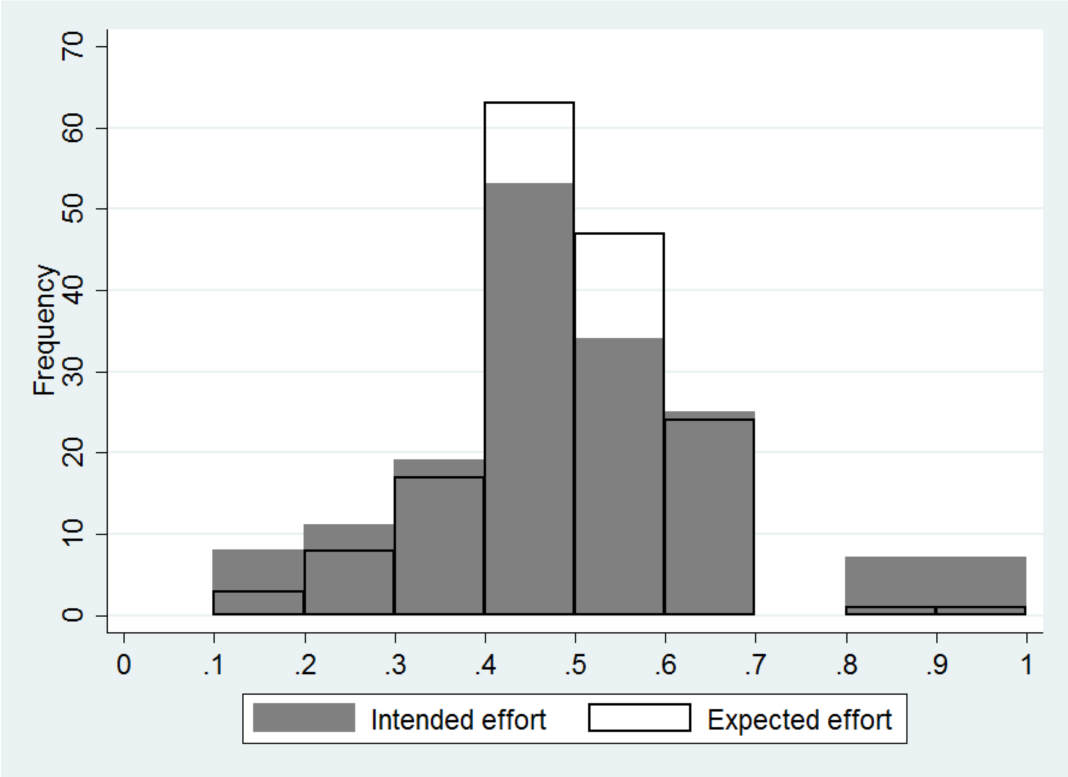
Treatment	$b_F^1 = e_N$	$e^0 = e_N$	$e = e_N$
Neutral	35.7%	32.1%	21.4%
Unkind	42.4%	31.8%	24.2%
Kind	35.7%	33.3%	16.7%

Table 3.3 shows, for each treatment, that over one third of the firms set their expectation of the workers' efforts, b_F^1, to be identical to the effort norm, e_N. A similar fraction of workers choose their intended effort, e^0, equal to the effort norm. Actual effort level, e, equals the norm in about a fifth of the cases. Figure 3.2 shows that expected effort by the firm (b_F^1) and the intended effort by the worker (e^0) are closely clustered around the effort norm $e_N = 0.4$. Since the norm is the actual average effort level in past experiments, these results have the following

[56] The data and results for each treatment are available from the authors on request.

[57] The *anchoring heuristic*, introduced by Daniel Kahenman and Amos Tversky is one of the most robust heuristics and widely confirmed by the evidence (Dhami, 2016; Section 19.6).

Figure 3.2: Histogram of expected and intended effort ($e_N = 0.4$).



interpretation. The announcement of an initial effort norm acts as a valuable anchor but the actual effort choice depends largely on the economic variables that are endogenous to the problem, such as the wage level, reciprocity and guilt.

Although we do not introduce an announced norm for wages, there might be a natural wage norm in these situations. To test this, we investigated if worker's first order belief (b_W^1) could be approximated well by the actual wage distribution (w). The nonparametric Mann-Whitney U test shows that these two distributions are not significantly different (Full data: $p = 0.322$; Neutral: $p = 0.3$; Unkind: $p = 0.49$; Kind: $p = 0.942$). The close connection between the wage expected by the worker and the actual wag suggests that we cannot rule out the existence of some norm for wages in our model. Furthermore, we compared the distribution of b_W^1 with the actual wage data from other studies[58]. Using the average wage level (42% of the maximum wage level) in Fehr et al. (1993) and employing a t-test[59], there is no significant difference between b_W^1 and w (Full data: $p = 0.182$; Neutral: $p = 0.486$; Unkind: $p = 0.337$; Kind: $p = 0.527$).

[58]The results from Fehr et al. (1993) are particularly clean. Fehr et al. (1998) reported the average wage was around 30%-65% of the maximum wage level in different treatments; Charness et al. (2004) reported the average wage was around 40%-50% of the maximum wage level in different treatments.

[59]Since we do not have the whole data of Fehr et al. (1993), we cannot do the Mann-Whitney U test here.

3.6.2 Testing the predictions of psychological game theory

In this section, we test the predictions of psychological game theory that are given in Propositions 3.1, 3.2 and Corollaries 1, 2. Note from Figure 3.1 that firms who offer higher wages also expect the workers to put in greater effort. Insofar as workers are informed of the expectations of firms (this follows from our induced beliefs methodology), workers who are made higher wage offers also have higher second order beliefs of the effort level that the firm expects them to put in. Hence, guilt aversion may drive them to put in more effort. This extra channel is missing in the sociological/anthropological explanations.

3.6.2.1 Negative belief-based reciprocity ($w < \frac{1}{2}\overline{w}$)

The predictions in this case are given in Proposition 3.1 and Corollary 1 and the predicted optimal effort lies in the interval $[0, b_W^2]$. We report, in Table 3.4, the percentage of cases in which the following three possible outcomes occur when all three treatments are combined– $e > b_W^2$, $e = b_W^2$, $e < b_W^2$. Overall, a little over four fifths of the subjects behave in a manner that is consistent with the predictions.

Table 3.4: Summary of effort levels in the domain of negative belief-based reciprocity.

Treatment	Observations	$e > b_W^2$	$e = b_W^2$	$e < b_W^2$
Combined	82	18.3%	22.0%	59.7%

The nonparametric Spearman rank correlation coefficient, r_s, between effort and wage in the three treatments is: Neutral ($r_s = 0.64$, $p = 0.006$), Unkind ($r_s = 0.37$, $p = 0.076$), Kind ($r_s = 0.71$, $p = 0.050$).[60] Thus, all treatments have a strong and significant positive correlations between effort and wage. This is consistent with theory (but see Remark 3.1).

Proposition 3.1 also implies that there should be no relation between the first order beliefs of the worker (b_W^1) and the effort level. The Spearman correlation coefficient, r_s, between e and b_W^1 in the three treatments are: Neutral ($r_s = 0.11$, $p = 0.563$), Unkind ($r_s = 0.37$, $p = 0.033$), Kind ($r_s = 0.08$, $p = 0.747$). Hence, none of the three treatments have significant correlations between effort and first order beliefs at the 1% level; this result is consistent with the theoretical predictions.

Corollary 1 shows that there is a positive correlation between second order beliefs, b_W^2, and effort, e. The Spearman rank correlation coefficient when data is pooled across all treatments, in the case $w < \frac{1}{2}\overline{w}$, is positive ($r_s = 0.25$) and significant at 5% ($p = 0.024$).

[60] The null hypothesis is that $r_s = 0$ and the alternative hypothesis is $r_s > 0$.

3.6.2.2 Positive belief-based reciprocity ($w > \frac{1}{2}\overline{w}$)

The predictions in this case are given in Proposition 3.2 and Corollary 2.

In the domain of positive reciprocity ($w > \frac{1}{2}\overline{w}$), all three results, $e > b_W^2$, $e = b_W^2$ and $e < b_W^2$ are possible. The higher is the reciprocity parameter, Y_R, the higher is optimal effort. Table 3.5 summarizes the relevant empirical results when data is combined across all treatments. Only 7.4% of the subjects exhibit $e > b_W^2$.

Table 3.5: Summary of effort levels in the domain of positive belief-based reciprocity.

Treatment	Observations	$e > b_W^2$	$e = b_W^2$	$e < b_W^2$
Combined	27	7.4%	40.7%	51.9%

The Spearman rank correlation between worker's second order belief and the (actual) effort is *not* significantly positive at the 1% level in any of the treatments. The results are as follows: Neutral ($r_s = 0.75$, $p = 0.013$), Unkind ($r_s = -0.04$, $p = 0.901$), Kind ($r_s = 0.67$, $p = 0.148$). Pooled across all treatments we have $r_s = 0.26$, $p = 0.209$ which is also not significantly positive. This is consistent with the sign prediction in Corollary 2 although it is not significant.

Proposition 3.2 implies that there should be no correlation between the first order beliefs of the worker (b_W^1) and the effort level, e. The Spearman correlation coefficient between e and b_W^1 in the three treatments is not significant at the 1% level in any treatment: Neutral ($r_s = 0.27$, $p = 0.445$), Unkind ($r_s = 0.43$, $p = 0.183$), Kind ($r_s = -0.25$, $p = 0.633$). This result is consistent with Proposition 3.2.

In the domain $w > \frac{1}{2}\overline{w}$, the predicted effect of wage level on the effort level is ambiguous when we restrict effort to $e \in (0, \overline{e}) / \{b_W^2\}$ (Proposition 3.2(v)). From Remark 3.1, we know that the only empirical case allowed by our data predicts a positive effect of wages on effort in this domain. The Spearman correlation coefficient between e and w when $e \in (0, \overline{e}) / \{b_W^2\}$ is positive ($r_s = 0.20$) but not significant ($p = 0.461$).

3.6.2.3 Summary

We may summarize the predictions of psychological game theory as follows. First, empirically effort is increasing in wage in the domain of negative belief-based reciprocity which is consistent with the predictions (but see Remark 3.1)); it has the right sign in the domain of positive belief-based reciprocity but it is not significant. Second, as predicted, the first order beliefs of the worker do not influence effort in the domains of positive and negative belief-based reciprocity. Third, the correlation of second order beliefs with effort is positive in the domains of negative and positive belief-based reciprocity (as predicted) but it is significant (at 5%) only in the domain of negative reciprocity. Overall, the predictions of psychological game theory are relatively more successful in the domain of negative reciprocity.

Remark 3.2. *: We are testing a version of the psychological game theory model in which the equitable payoff in (3.7) combines the minimum and maximum payoffs in the proportions* $0.5 : 0.5$*; in this we follow the existing literature. This gives rise to the domains of negative belief-based reciprocity (*$w < \frac{1}{2}\overline{w}$*) and positive belief-based reciprocity (*$w > \frac{1}{2}\overline{w}$*). In actual practice social norms that vary across different cultures might combine the maximum and minimum payoffs in the proportions* $\gamma : 1-\gamma$, $\gamma \in [0,1]$*. In this, more general case, the appropriate domains of negative and positive belief-based reciprocity are* $w < \gamma\overline{w}$ *and* $w > \gamma\overline{w}$*, respectively. It is a-priori not known what value of* γ *to use for the subjects in our experiments. A mis-specification of* γ *may create domains over which the predictions of PGT are be unfairly evaluated. Since* γ *is unobserved in our study, and in the literature, this remains a shortcoming of our work. It might be more illuminating to consider the Spearman correlation between effort and second order beliefs across the entire sample. This is positive and significant in all treatments: Neutral (*$r_s = 0.43$ *,* $p = 0.001$*), Unkind (*$r_s = 0.33$*,* $p = 0.007$*), Kind (*$r_s = 0.36$*,* $p = 0.020$*). Across all treatments we have* $r_s = 0.37$*,* $p = 0.000$*, which is significantly positive, suggesting the presence of simple guilt–aversion.*

3.6.3 Testing the predictions of Akerlof's gift exchange formulation

We now test the predictions of Akerlof's formulation of the gift exchange game (see Proposition 3.3).

In line with the prediction of Proposition 3.3(i), in the case of $w > 0$, the effort levels chosen by around 73.29% of the subjects are strictly greater than the minimum effort level, 0.1.

Proposition 3.3(ii) predicts that effort is positively correlated with wage when $e^* \in (0, \overline{e})$[61]. The Spearman correlation coefficient supports the prediction in the Neutral treatment ($r_s = 0.58$, $p = 0.000$) and the Unkind treatment ($r_s = 0.44$, $p = 0.000$). In the Kind treatment, the Spearman correlation is positive ($r_s = 0.14$) but not significant ($p = 0.394$). Therefore, on the whole, our experimental results corroborate Proposition 3.3(ii).

[61] In the experimental setting, the lowest effort level is 0.1, which is equivalent to 0 in our theoretical section. Hence, the experimental data for effort are restrained to $(0.1, 1)$ here.

Proposition 3.3(ii) predicts that the beliefs of the players do not influence the effort level. This is not borne out by the data. We have already considered the significantly positive correlation between effort and second order beliefs in Remark 3.2. We also compute the Spearman correlation of effort with first order beliefs. It is not significantly different from zero in the Neutral and Kind treatments but we cannot reject that it is zero at the 1% level in the Unkind treatment: Neutral ($r_s = 0.04$, $p = 0.790$), Unkind ($r_s = 0.33$, $p = 0.007$), Kind ($r_s = 0.01$, $p = 0.970$). Across all treatments we have $r_s = 0.14$, $p = 0.070$.

Summary: The Akerlof's gift exchange model predicts well the effects of wages on effort but it does not predict the effect of beliefs on effort, which is predicted in psychological game theory.

3.6.4 Testing the MS formulation of gift exchange

The predictions for the MS formulation of gift exchange are given in Proposition 3.4. From Proposition 3.4(i), the theory predicts that in the domain of negative gift exchange ($w < b_W^1$), effort should take the lowest possible value, which in our experiments is $e^* = 0.1$. However, while 37.5% of the subjects put in an effort level of 0.1, the effort levels of others are distributed over higher effort levels; 17.2% of the subjects put in an effort level greater than or equal to 0.4.

Table 3.6: Spearman rank correlations of effort with wage and first order beliefs.

Treatment	$Corr(e^*, w)$	$Corr(e^*, b_W^1)$
Neutral	0.22 (0.316)	0.35 (0.106)
Unkind	0.3 (0.123)	0.55 (0.003)
Kind	0.03 (0.897)	0 (0.992)

Proposition 3.4(ii) predicts that in the domain of positive gift exchange ($w > b_W^1$), all workers should exert *nonzero* effort level (in our experiments, this corresponds to $e > 0.1$). This is confirmed in all treatments. Furthermore, Proposition 3.4(ii) and Remark 3.1 also predict that: (1) The effect of wage, w, on effort is positive. (2) Effort (e) and first order belief (b_W^1) should be negatively correlated. The Spearman rank correlations in each case, $Corr(e^*, w)$ and $Corr(e^*, b_W^1)$, respectively, are shown in Table 3.6; the numbers in the parentheses are the p-values. $Corr(e^*, w)$ is positive but we cannot reject the null hypothesis of zero correlation at the 10% level. $Corr(e^*, b_W^1)$ is positive (not *negative* as predicted) but it is significant (at 1%) only at the Unkind treatment.

Summary: There is mixed support for the MS formulation. The effort choice under positive gift exchange has better conformity with the evidence relative to the case of negative gift exchange. While the Spearman correlations of effort with wage are generally of the correct sign but most coefficients are not significant; the correlations of effort with first order beliefs are of the opposite sign. The model does not give any role to second order beliefs in determining effort which are significant for the entire sample (Remark 3.2).

3.6.5 Model comparison tests

The experimental support for the three competing models of gift exchange is impressive, although psychological game theory (PGT) and Akerlof's formulation appear to fare better than the MS formulation of gift exchange. Over the entire domain, there is a significant and positive correlation of effort with second order beliefs, but this is not predicted by Akerlof's gift exchange model. By contrast, Akerlof's model successfully predicts the effect of wages on effort, while under PGT the effect is generally ambiguous and it depends on the values of the parameters (in our case the effect is positive). The PGT model has more parameters than Akerlof's model, so one needs to check formally if this can be traded off against the potentially greater explanatory power of the PGT models in explaining the role of beliefs. In this section, we use statistical techniques to select the best-fitting model that takes accounts of such tradeoffs.

Table 3.7, shows the three robust OLS regression models.[62] We obtained the demographic characteristics of the subjects using the post-experimental survey. None of the demographic variables had a significant effect on effort and had little effect on the significance of the other variables.

Model 1 used the effort data in the domain $e \in (0.1, 1)$, and tested the PGT model. In Model 1, wage significantly and positively affects effort, which is consistent with the prediction of the PGT model for the parameter values and choices made by subjects, although, in general, the prediction is ambiguous (Propositions 3.1, 3.2 and Remark 3.1). Second order beliefs do not influence effort significantly, but still take the expected positive sign (to reflect guilt-aversion) and the magnitude is large. The only significant explanatory variable is the wage level.

Model 2 used the effort data in the domain of $(0.1, 1)$ to test Akerlof's gift exchange model (from Proposition 3.3, there is no variability with respect to the exogenous

[62] In our case, 'robust' refers to clustering on individuals (individual subjects), i.e., observations for individual i are correlated in some unknown way with his own unobserved characteristics, but the errors across individuals i and $j \neq i$ are not correlated.

Table 3.7: Regressions results.

Dependent Variable	Effort		
Data	$e \in (0.1, 1)$	$e \in (0.1, 1)$	$w > b^1_W$
Model	Psychological	Akerlof	MS
	1	2	3
Wage	0.003**	0.005***	0.003
	[0.001]	[0.001]	[0.003]
FOB			0.004*
			[0.002]
SOB	0.257		
	[0.189]		
D_U	0.013	0.021	−0.032
	[0.025]	[0.027]	[0.051]
D_K	0.037	0.055	−0.010
	[0.031]	[0.033]	[0.073]
Constant	0.146**	0.186***	0.041
	[0.066]	[0.052]	[0.052]
Observations	116	116	70
F	7.11***	13.14***	4.02***
Adjusted R^2	0.183	0.164	0.110

Note: The treatment dummy $D_U = 1$ for Unkind treatment, and 0 otherwise; $D_K = 1$ for Kind treatment, and 0 otherwise. We show robust standard errors in the brackets; *** denotes the case $p < 1\%$, ** denotes $p < 5\%$, and * denotes $p < 10\%$.

variables at the corner solution). As in Model 1, effort increases significantly in the wage, which is consistent with the prediction of Proposition 3.3. Akerlof's gift exchange model makes no predictions about the role of beliefs, so these are not included in the regression. Notably, beliefs lose significance in the presence of wages (Model 1).

Model 3 only used the data in the domain of positive gift exchange ($w > b^1_W$) to test the MS formulation of gift exchange; see Proposition 3.4, which shows that effort is zero for negative gift exchange ($w < b^1_W$). The wage effect on effort is not significantly different from zero although for the parameter values, the prediction is that effort should be increasing in wages (Proposition 3.4, Remark 3.1). First order beliefs positively and significantly influence effort; this is inconsistent with Proposition 3.4, which predicts a negative effect.

Since Models 1 and 2 are *nested* models, we can directly compare the adjusted R^2 to determine their explanatory powers. The PGT model and the Akerlof model have similar adjusted R^2, and both are higher than the MS gift exchange model.

Second order beliefs do not appear to significantly explain effort choices, because the adjusted R^2 increases only slightly as we move from Model 2 to Model 1.

To consider the tradeoff between greater explanatory power and a larger number of parameters in Model 1 relative to Model 2, we also report the Akaike and the Bayesian information criteria (AIC and BIC) in Table 3.8.

Table 3.8: Model comparison.

Model	PGT	Akerlof	MS
Adjusted R^2	0.015	0.031	0.011
AIC	-65.515	-67.247	-65.297
BIC	-55.856	-59.520	-55.638

In computing the statistics in Table 3.8, we have used data that is common to all models ($e \in (0.1, 1)$ and $w > b_W^1$).[63] On these criteria, Akerlof's gift exchange model is the best because of the lowest AIC/BIC (parsimony) and the highest adjusted R^2 (fit). We are not arguing these statistical criteria are the sole or even the most important criteria for choosing among models. One might be more interested in how beliefs determine outcomes or how well the models perform in different domains– these economic reasons for choice may supersede purely statistical criteria.

3.7 Conclusions

We pit alternative explanations for the behavior of workers in gift exchange experiments against each other. The three main models that we focus on are Akerlof's formulation, MS formulation of gift exchange, and a model based on psychological game theory. The first two models lie within the ambit of sociological/anthropological theories that stress action-based reciprocity as the main driving force behind the observed phenomenon of workers putting in higher effort in response to higher wages. Models of psychological game theory stress the role of intentions and, in particular, the roles of belief-based reciprocity and guilt aversion in influencing human behavior. We derive the theoretical predictions of each class of models and then test them with a fresh dataset. We use the induced beliefs method which, as far as we know, has not been used for the gift exchange game in the past.

We find that the experimental support for the psychological game theory (PGT) model and for Akerlof's formulation appear to be stronger than for the MS formulation of gift exchange. There is a significant and positive correlation of effort with second order beliefs. This is predicted by the PGT model but is ruled out by the

[63] As mentioned above, Models 1 and 2 used the effort data in the domain of $(0.1, 1)$ and Model 3 used data from the domain $w > b_W^1$.

other two models. Akerlof's model predicts a positive effect of wages on effort. This is strongly supported by the evidence. The PGT predicts the same but only for certain parameter ranges.

When we use model selection tests that allow for a trade-off between extra parameters and greater explanatory power, we find that Akerlof's gift exchange formulation is the clear winner. We conjecture that the cognitive simplicity involved in following a simple norm of action-based reciprocity may be the prime explanation behind the relative success of Akerlof's formulation.

Chapter 4

Can quantum decision theory explain the Ellsberg paradox?

Mengxing Wei, Ali al-Nowaihi, Sanjit Dhami

Abstract

We consider a simple quantum decision model of the Ellsberg paradox. We report the results of an experiment we performed to test the matching probabilities predicted by this model using an incentive compatible method. We find that the theoretical predictions of the model are in conformity with our experimental results. This supports the thesis that violations of classical (Kolmogorov) probability theory may not be due to irrational behaviour but, rather, due to inadequacy of classical probability theory for the description of human behaviour. Unlike earlier quantum models of the Ellsberg paradox, this model makes essential use of quantum probability and gives a parameter-free derivation of the matching probabilities.

4.1 Introduction

Situations of ambiguity are pervasive in decision making. The most successful approach is probably that of *source dependence* (Abdellaoui et al., 2011; Kothiyal et al., 2014; Dimmock et al., 2015).[64] In this paper, we investigate the potential of *quantum decision theory* (QDT) to provide an alternative explanation. We concentrate on the canonical example of ambiguity, namely, the Ellsberg paradox (Keynes, 1921; Ellsberg 1961, 2001). The Ellsberg paradox has proved to be a particularly useful vehicle for testing models of ambiguity.

Consider the following version of the Ellsberg experiment due to Dimmock et al. (2015). This involves two urns: The known urn (K) contains kn balls of n different colors and k balls of each color. The unknown urn (U) also contains kn balls of the same n colors as urn K but in unknown proportions. The subject is presented with the following bet. Suppose l of the n colors are chosen to be winning colors (hence, urn K contains kl balls of the winning colors). The subject wins a prize if a randomly drawn ball from an urn is of the winning color. Which of the two urns (K or U) should the subject choose?

By the heuristic of *insufficient reason* (or *equal a-priori probabilities*)[65], the probability of drawing a ball of a winning color out of urn K is $p = \frac{kl}{kn} = \frac{l}{n}$. Although experimental subjects do not know the proportions of the different colors in urn U, they have no reason to favour one proportion over another. Hence, by the heuristic of insufficient reason, they should assign the same probability, $p = \frac{l}{n}$, to drawing a ball of a winning color from urn U. It follows that they should have no reason to prefer K to U or U to K on probabilistic grounds. They should be *ambiguity neutral.* However, what is observed in Dimmock et al. (2015) is the following. Subjects prefer U for low p but K for high p, i.e., they are *ambiguity seeking* for low probabilities but *ambiguity averse* for high probabilities.[66] This behaviour is called *Insensitivity.* Thus classical theory predicts ambiguity neutrality while observation supports insensitivity. We shall call this behavior the *Ellsberg paradox.*[67]

Consider subject i. Keep the contents of urn U fixed, but construct a new known

[64] See section 3 of al-Nowaihi and Dhami (2017) for a review of classical (non-quantum) approaches to ambiguity.

[65] *Insufficient reason* or *equal a-priori probabilities* is now commonly referred to as *indifference.* However, indifference has a well-established alternative meaning in economics. To avoid confusion, we shall use the older terminology.

[66] This terminology is in analogy to situations of risk, where a decision maker is risk averse (risk neutral, risk loving) if the certainty equivalent of a lottery is less (equal to, greater) than the expected value.

[67] Traditionally, the Ellsberg paradox is used to refer to ambiguity aversion only. Our usage is in conformity with Ellsberg's original usage (see Ellsberg, 2001) and recent scholarship (see Dimmock, et al., 2015).

urn, K_i, with a known number, M_i, of balls of the winning colors such that subject i is indifferent between urns K_i and U. Then $m_i(p) = \frac{M_i}{kn} = \frac{M_i}{k(l/p)}$ is the *matching probability* of p for subject i. Note that the definition of $m_i(p)$ is operational and does not depend on the particular decision theory assumed for the subjects. Let there be N subjects and let $m(p) = \frac{1}{N}\sum_{i=1}^{N} m_i(p)$ be the average of matching probabilities across all subjects. In their empirical exercise, Dimmock et al. (2015) report $m(0.1) = 0.22$, $m(0.5) = 0.40$, $m(0.9) = 0.69$. Thus, on average, subjects are ambiguity seeking for low probabilities ($m(0.1) > 0.1$) but ambiguity averse for medium and high probabilities ($m(0.5) < 0.5$, $m(0.9) < 0.9$).

A number of quantum models of the Ellsberg paradox have been developed.[68] Some of these models predict universal ambiguity aversion or universal ambiguity seeking, thus are in conflict with the observed insensitivity. Others do explain the Ellsberg paradox but at the cost of introducing a certain degree of flexibility. However, when non-quantum models are granted the same degree of flexibility, they too can explain the Ellsberg paradox. Busemeyer and Bruza (2012, section 9.1.2) conclude "In short, quantum models of decision making can accommodate the Allais and Ellsberg paradoxes. But so can non-additive weighted utility models, and so these paradoxes do not point to any unique advantage for the quantum model". Furthermore, there is considerable arbitrariness in the choice of weights in weighted utility models. Hence they introduce flexibility at the cost of lower predictive power.

A simple quantum model of the Ellsberg paradox was introduced by al-Nowaihi and Dhami (2017). They replace weights with quantum probabilities which are parameter-free. Thus, their explanation of the Ellsberg paradox is more parsimonious, hence more refutable, than all the other explanations. Their predicted matching probabilities, based on their quantum model, are $m(0.1) = 0.171$, $m(0.5) = 0.417$, $m(0.9) = 0.695$, which are close to those empirically observed by Dimmock et al. (2015). The three main assumptions they employ are:

1. (Q) Quantum probability theory.

2. (I) The heuristic of insufficient reason.

3. (B) A behavioral assumption on how urn U is constructed in a subject's mind.

We add a fourth assumption to the above for the purpose of testing the theory:

4. (P) A power function form for the utility function.

[68] See the Introduction of al-Nowaihi and Dhami (2017) for a review of earlier quantum approaches to the Ellsberg paradox.

The first assumption (Q) is the main assumption. However, no mathematical structure on its own will yield empirically testable predictions; auxiliary assumptions are needed.[69] In this respect, al-Nowaihi and Dhami (2017) is no exception. In addition to the assumptions of quantum probability theory, they employ the two auxiliary assumptions (I) and (B). These three assumptions are sufficient to theoretically derive the matching probabilities (Proposition 4.2, section 4.5). Proposition 4.1, section 4.3, shows that combining the auxiliary assumptions (I) and (B) with standard (Kolmogorov) probability theory produces ambiguity neutrality, contrary to the evidence for insensitivity. Thus, al-Nowaihi and Dhami (2017) make essential use of quantum probability theory.

Testing any theory requires further assumptions.[70] We have added assumption (P) for the purpose of testing the theory. Thus, our test is a test of the conjunction Q&I&B&P. If we reject this conjunction, then this is a rejection of, at least, one of them, but we would not know which.[71] However, since Q&I&B&P is true if, and only if, all of these are true, then a confirmation is a confirmation of each one of them. However, a confirmation is not a proof. It is merely a failure to reject. Hence further tests may lead to a rejection.[72]

We now turn to a further discussion of assumptions 1-4.

4.1.1 Quantum probability theory (Q)

Quantum decision theory (QDT) originated with Aerts and Aerts (1994) who noticed similarities between paradoxes of human behaviour and paradoxes of quantum mechanics.

From a classical point of view, the results of quantum mechanics appear paradoxical. This led von Neumann (1955), to devise a new mathematical structure in which quantum mechanics can be given a consistent formulation, *Hilbert Space* and *quantum probability.* Events are vector subspaces of Hilbert space, and quantum probability is an additive (though not distributive) measure on these.

In *quantum decision theory* (QDT), unlike all other decision theories, events are

[69]For example, in Newtonian mechanics, in addition to Newton's second law of motion and law of gravity, we need initial conditions and simplifying assumptions. Calculus on its own will not yield empirically testable predictions. In quantum mechanics we need, for example, the momentum operator to be $p_x = -i\frac{h}{2\pi}\frac{\partial}{\partial x}$ and we need to specify a Hamiltonian for the system. Hilbert space on its own is insufficient.

[70]For example, to test Newton's prediction of the orbits of the planets we need to make assumptions about the human eye, the telescope and the atmosphere.

[71]The first test of Maxwell's electromagnetic theory, by Hertz in 1888, led to a rejection. However, Hertz conjectured that it was one of the auxiliary assumptions that was rejected, not Maxwell's theory. Hertz's conjecture was later confirmed. See Chalmers (1999, pp. 31-35).

[72]No number of confirmations, however large, can prove a theory. The most we can say about a theory, any theory, is that it has so far survived the tests.

not distributive, and this is the main difference between the two. Thus, in QDT the event "X *and* (Y *or* Z)" need not be equivalent to the event "(X *and* Y) *or* (X *and* Z)". On the other hand, in all other decision theories, these two events are equivalent. This non-distributive nature of QDT is the key to its success in explaining paradoxes of behaviour that other decision theories find difficult to explain. For example, *order effects*, the *Linda paradox*, the *disjunction fallacy* and the *conjunction fallacy*.[73] As a result of the non-distributive nature of QDT, the *law of total probability* does not generally hold. Instead, we use the *Feynman rules* and the *law of reciprocity*.[74] We refer the reader to al-Nowaihi and Dhami (2017, section 4) for an introduction to the quantum concepts and tools needed for this paper. For an excellent book-length introduction to quantum decision theory, see Busemeyer and Bruza (2012). For papers examining the limits of standard quantum theory when applied to cognitive psychology, see Khrennikov et al. (2014), Basieva and Khrennikov (2015) and Asano (2016).

One can take either of the following two positions:

1. Rational beings should follow Kolmogorov probability theory. The more general quantum probability would then give a systematic account of irrational human behaviour.

2. Kolmogorov probability theory is simply inadequate to describe human behaviour. We need a more general probability theory, such as quantum probability or Choquet capacity.

The situation is analogous to the following attitudes to the St. Petersburg paradox:

1. Rational beings should follow expected value theory. The more general expected utility theory would then give a systematic account of irrational human behaviour.

2. Expected value theory is simply inadequate to describe human behaviour. We need a more general theory such as expected utility theory.

We prefer the second position.

4.1.2 The heuristic of insufficient reason (I)

In both classical (Kolmogorov) probability theory and quantum probability theory any probabilities (provided they are non-negative and sum to 1) can be assigned to the elementary events. To make a theory predictive, some heuristic rule is needed to assign a' priori probabilities (we call this a heuristic because it does not follow from

[73] See Busemeyer and Bruza (2012). In particular, sections 1.2, 4.1-4.3, 5.2 and 10.2.3.

[74] See Busemeyer and Bruza (2012), pp. 5, 13, 39.

either classical or quantum probability theory). The heuristic commonly used is that of insufficient reason or equal a' priori probabilities.[75] This heuristic is crucial in deriving the Maxwell-Boltzmann distribution in classical statistical mechanics and the Bose-Einstein and Fermi-Dirac distributions in quantum statistical mechanics.[76]

4.1.3 A behavioral assumption on how urn U is constructed in a subject's mind (B)

The framing of information is vital in choices. Subjects often simplify complex problems before solving them (Dhami, 2016)[77]. For Ellsberg experiments, subjects are typically told that urn U contains the same number of balls of the same colors as urn K, but in unknown proportions. However, the term "unknown proportions" is not defined any further, which raises the question of how subjects perceive this term. Pulford and Colman (2008) provide strong evidence that this is too cognitively challenging for subjects and that subjects do not consider all possible distributions of balls in urn U. al-Nowaihi and Dhami (2017) introduce a simple assumption on how subjects construct urn U in their mind (see section 4.3, below). As discussed above, confirmation of the model is a confirmation of the conjunction Q&I&B&P and, in particular, a confirmation of (B). But a confirmation is not a proof. Further tests are always required. One such test is to test (B) independently. But we leave this for future research.

4.1.4 A power function form for the utility function (P)

The matching probabilities by al-Nowaihi and Dhami (2017) are close to those empirically observed by Dimmock et al. (2015). However, the mechanism used by Dimmock et al. (2015) is not incentive compatible. Specifically, Dimmock et al. (2015) constructed urn K_i as follows. The ratio of the colors (whatever they are) in U were kept fixed. However, the ratio in K_i was varied until subject i declared indifference between K_i and U. It turns out that in this method of eliciting matching probabilities subjects have the incentive to declare a preference for U over K_i, even when the reverse is true. However, Dimmock et al. (2015) found no evidence

[75]To be sure, this heuristic is not without problems. See, for example, Gnedenko (1968), sections 5 and 6, pp 37 to 52.

[76]See Tolman, 1938, section 23, pp 59-62, for a good early discussion.

[77]Examples and analysis are provided throughout Dhami (2016); see, for instance, Part 7.

in their data that this occurred.[78]

In this paper, we report the results of an experiment we performed using a new data set and the incentive compatible mechanism of Fox and Tversky (1995, study 2). However, the Fox and Tversky (1995) method requires the elicitation of the subjects' utility functions (this is not required by the Dimmock et al., 2015, mechanism). In turn, this requires the specification of a utility function, u_i, for each subject, i. As in Fox and Tversky (1995, study 2) we use the power function form $u_i(x) = x^{\sigma_i}, x \geq 0, \sigma_i > 0$.[79] This introduces a free parameter, σ_i. Note, however, that σ_i is only used to give a parsimonious description of the behaviour of subjects (see sections 4.7, below). In particular, σ_i is not chosen to make the predictions of the theory fit the evidence. The matching probabilities predicted by the theory are parameter-free and are based on assumptions (Q), (I) and (B) only (see Proposition 4.2, section 4.5).

4.1.5 Organization of the paper

The rest of the paper is organized as follows. Section 4.2 gives the main stylized facts from Ellsberg experiments. Section 4.3 gives the main behavioral assumption of al-Nowaihi and Dhami (2017). Proposition 4.1 of section 4.5 shows that the Ellsberg paradox reemerges when this behavioral assumption is combined with classical (Kolmogorov) probability theory. Hence al-Nowaihi and Dhami (2017) makes essential use of quantum probability theory. Proposition 4.2 of section 4.5 gives the main theoretical predictions of al-Nowaihi and Dhami (2017). Section 4.6 gives our experimental design. Sections 4.7 and 4.8 give our experimental results. Section 4.9 summarizes and concludes. Appendix C.1 gives our experimental instructions. Appendix C.2 gives our post-experimental questionnaire.

4.2 Stylized facts

The following are the main stylized facts of Ellsberg experiments.

1. *Insensitivity*: Subjects are ambiguity averse for medium and high probabilities but ambiguity seeking for low probabilities; see Dimmock et al. (2015) for a

[78] Dimmock et al. (2015, p26): "In chained questions, where answers to some questions determine subsequent questions, subjects may answer strategically (Harrison, 1986). In our experiment, this is unlikely. First, our subjects are less sophisticated than students. Second, it would primarily have happened in the end (only after discovery), at the 0.9 probability event, where it would increase ambiguity seeking. However, here we found strong ambiguity aversion".

[79] The power form of the utility function is a popular choice (Kahneman and Tversky, 2000). In particular, see Benartzi and Thaler (1995), Fox and Tversky (1995), Prelec (1998), Thaler (1999) and Tversky and Kahneman (1992). For an axiomatization, see al-Nowaihi et al. (2008).

recent survey of the literature as well as their own experimental results.

2. *Exchangeability*: Subjects are indifferent between colors. Subjects are indifferent between being asked to choose a color first or an urn first (Abdellaoui et al., 2011).[80]

3. *No error*: Suppose that a subject prefers one urn (K or U) over the other. It is then explained to the subject that, according to classical probability theory, she should have been indifferent. She is offered the chance to revise her assessment. Subjects usually decline to change their assessment (Curley et al., 1986).

4. *Salience*: Ambiguity aversion is stronger when the two urns are presented together than when they are presented separately (Fox and Tversky, 1995; Chow and Sarin, 2001, 2002).[81]

5. *Anonymity* (or *fear of negative evaluation*): Ambiguity aversion does not occur if subjects are assured that their choice between urn U and urn K is anonymous (Curley et al., 1986; Trautmann et al., 2008).

In this paper, we show that the model of al-Nowaihi and Dhami (2017) is in accord with stylized fact 1, *insensitivity*, both qualitatively and quantitatively. It is also in accord with stylized facts 2 (*exchangeability*) and 3 (*No error*).

It may also be in accord with stylized facts 4 (*salience*) and 5 (*anonymity*). Suppose l of the n colors are winning colors. If a subject is presented with the two urns separately, or if the choice is made anonymously, then, maybe, that subject simply uses the heuristic of insufficient reason to conclude that the probability of drawing a winning ball is $\frac{l}{n}$, whether the subject is choosing from urn K or urn U. However, if the subject is presented with urns K and U together, and the choice is not under anonymity, then the subject may feel compelled to reason it through. However, the detailed development and testing of this is beyond the scope of this paper.

[80]As an example, suppose that there are 100 balls each in urn K and urn U. There are two colors in Urn K, black and white (so 50 balls of each color). In contrast, in Urn U, the two colors (black and white) are in unknown proportion. The subject is told that if a ball of the color of his choice is drawn from an urn of his choice, then he wins a prize $z. We may now proceed in one of two alternative ways. (1) First ask the subject to choose a color, then an urn. (2) First ask the subject to choose an urn, then a color. Exchangeability requires the answers in the two methods to be identical.

[81]Even more strikingly, Fox and Tversky (1995) found that for probability $\frac{1}{2}$, subjects exhibited ambiguity aversion with the value of urn U remaining approximately the same but urn K revalued upwards. Chow and Sarin (2001, 2002) did not find this result, but did find that ambiguity aversion is more pronounced when subjects are presented with K and U together.

4.3 Behavioral assumption: Construction of urn U in the mind of a subject

Recall, from the Introduction, that the known urn (K) contains kn balls of n different colors and k balls of each color. The unknown urn (U) also contains kn balls of the same n colors as urn K but in unknown proportions. The subject is asked to select one of the urns (K or U). A ball is drawn at random from the urn chosen by the subject. Suppose that l of the n colors are the winning colors (hence, urn K contains kl balls of the winning colors).

al-Nowaihi and Dhami (2017) conjecture that subjects model "unknown proportions" in a simple way, as described below.[82]

1. We replace colors by numerals (this is justified by stylized fact 2). Furthermore, we consider only two numerals: 1 and 2. The known urn K contains kn balls, kl of which are labeled "1" and $kn - kl$ are labeled "2". We shall adopt the heuristic of insufficient reason. Thus, ball 1 is drawn from K with probability $p = \frac{kl}{kn} = \frac{l}{n}$ and ball 2 is drawn from K with probability $1-p = \frac{kn-kl}{kn} = \frac{n-l}{n}$.[83]

2. Point 1 allows us to consider urn K as simply having two balls – one of the balls, the winning ball labeled "1", is drawn with probability $p = \frac{l}{n}$. The only other remaining ball, labeled "2", is drawn with probability $1-p = \frac{n-l}{n}$. To compare with the evidence reported in Dimmock et al. (2015), we are interested in $p = 0.1$, 0.5 and 0.9. Likewise urn U will also have two balls labeled 1 and 2 but the proportions will be unknown, as the following construction shows.

3. A subject is presented with two urns, K and U. Urn K has two balls, labeled 1 and 2, while urn U is initially empty. We conjecture that in the mind of a subject urn U is constructed as follows. In two successive and independent rounds, a ball is drawn at random from urn K and placed in urn U without revealing the labels, 1 or 2, to the subject. At the end of each of the two rounds, the ball that was drawn from urn K is replaced with an identically labeled ball. At the end of the two rounds, urn U contains two balls. The possibilities are that both could be labeled 1, both could be labeled 2, or one could be labeled 1 and the other labeled 2.

4. A ball is drawn at random from whichever urn the subject chooses (K or U). The subject wins a monetary prize $v > 0$ if ball 1 is drawn but wins nothing

[82] Recall subsection 4.1.3 of the Introduction.

[83] This transformation is only for analytic convenience. In our experiments subjects are always presented with colored balls whose ratios match the probabilities.

if ball 2 is drawn.

5. Since we have two balls and two states, we work in a 4-dimensional space.

Based on the above construction, we may define the following states:

1. $\mathbf{s}_1$ is the state where ball 1 is drawn in each of the two rounds (each with probability p).
2. $\mathbf{s}_2$ is the state where ball 1 is drawn in round one (probability p) then ball 2 is drawn in round two (probability $1-p$).
3. $\mathbf{s}_3$ is the state where ball 2 is drawn in round one (probability $1-p$) then ball 1 is drawn in round two (probability p).
4. $\mathbf{s}_4$ is the state where ball 2 is drawn in each of the two rounds (each with probability $1-p$).

Our behavioral assumption about how a subject mentally constructs urn U, outlined above (in particular, point 3), will play an essential role in explaining the Ellsberg paradox. The question then arises whether this behavioral assumption can also explain the Ellsberg paradox when combined with classical (Kolmogorov) probability theory. Proposition 4.1, below, establishes that this is not the case.

Proposition 4.1. *: If the probability of drawing ball* 1 *from the known urn* K *is* p*, then the classical probability of drawing ball* 1 *from the unknown urn* U *is also* p*.*

Proof of Proposition 4.1: Let $\mathbf{t}$ be the state where ball 1 is drawn. By the *law of total probability*, we then have:

$$P(\mathbf{t}) = P(\mathbf{t}|\mathbf{s}_1)\,P(\mathbf{s}_1) + P(\mathbf{t}|\mathbf{s}_2)\,P(\mathbf{s}_2) + P(\mathbf{t}|\mathbf{s}_3)\,P(\mathbf{s}_3)\,. \tag{4.1}$$

We have $P(\mathbf{t}|\mathbf{s}_1) = 1$, $P(\mathbf{s}_1) = p^2$, $P(\mathbf{t}|\mathbf{s}_2) = \frac{1}{2}$, $P(\mathbf{s}_2) = p(1-p)$, $P(\mathbf{t}|\mathbf{s}_3) = \frac{1}{2}$, $P(\mathbf{s}_3) = (1-p)\,p$. Hence, from (4.1), we get:

$$P(\mathbf{t}) = p. \tag{4.2}$$

Hence, if the probability of drawing ball 1 from the known urn K is p, then the classical probability of drawing ball 1 from the unknown urn U is also p. ■

Thus, even with our behavioral assumption, the classical treatment gives the same probability, p, of winning whether a subject chooses urn K or urn U. Hence, if a subject strictly prefers K to U (or U to K), then this subject is violating classical theory.

This is in contrast to other quantum explanations of the Ellsberg paradox (see the Introduction of al-Nowaihi and Dhami, 2017, for a review). These other explanations introduce auxiliary assumptions that when combined with classical (non-quantum) probability theory can also explain the Ellsberg paradox.

4.4 Elements of quantum probability theory

4.4.1 Vectors

For our purposes (as we shall show), it is sufficient to use a finite dimensional real vector space $\mathbb{R}^n$ (in fact, with $n = 2$ or $n = 4$). A vector, $\mathbf{x} \in \mathbb{R}^n$, is represented by an $n \times 1$ matrix (n rows, one column). Its *transpose*, $\mathbf{x}\dagger$, is then the $1 \times n$ matrix (one row, n columns) of the same elements but written as a row.[84]. The *zero vector*, $\mathbf{0}$, is the vector all of whose components are zero. Let $r \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ with components x_i and y_i, respectively. Then $r\mathbf{x}$ is the vector whose components are rx_i and $\mathbf{x}+\mathbf{y}$ is the vector whose components are x_i+y_i. $\mathbf{y} \in \mathbb{R}^n$ is a *linear combination* of $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m \in \mathbb{R}^n$ if $\mathbf{y} = \sum_{i=1}^{m} r_i \mathbf{x}_i$ for some real numbers $r_1, r_2, ..., r_m$. The *inner product* of $\mathbf{x}$ and $\mathbf{y}$ is $\mathbf{x}\dagger\mathbf{y} = \sum_{i=1}^{n} x_i y_i$, where x_i, y_i are the components of $\mathbf{x}$ and $\mathbf{y}$, respectively.[85] If $\mathbf{x}\dagger\mathbf{y} = 0$, then $\mathbf{x}$ is said to be *orthogonal* to $\mathbf{y}$ and we write $\mathbf{x} \perp \mathbf{y}$. Note that $\mathbf{x} \perp \mathbf{y}$ if, and only if, $\mathbf{y} \perp \mathbf{x}$. The *norm*, or *length*, of $\mathbf{x}$ is $\|\mathbf{x}\| = \sqrt{\mathbf{x}\dagger\mathbf{x}}$. $\mathbf{x}$ is *normalized* if $\|\mathbf{x}\| = 1$.[86] $X \subset \mathbb{R}^n$ is a *vector subspace* (of $\mathbb{R}^n$) if it satisfies: $X \neq \emptyset$, $\mathbf{x}, \mathbf{y} \in X \Rightarrow \mathbf{x} + \mathbf{y} \in X$ and $r \in \mathbb{R}, \mathbf{x} \in X \Rightarrow r\mathbf{x} \in X$. Let $\mathcal{L}$ be the set of all vector subspaces of $\mathbb{R}^n$. Then $\{\mathbf{0}\}, \mathbb{R}^n \in \mathcal{L}$. Let $X, Y \in \mathcal{L}$. Then $X \cap Y \in \mathcal{L}$ and $X + Y = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in X, \mathbf{y} \in Y\} \in \mathcal{L}$. If $X_1, X_2, ..., X_m \in \mathcal{L}$, then $\sum_{i=1}^{m} X_i = \{\sum_{i=1}^{m} \mathbf{x}_i : \mathbf{x}_i \in X_i\} \in \mathcal{L}$. The *orthogonal complement* of $X \in \mathcal{L}$ is $X^{\perp} = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} \perp \mathbf{x} \text{ for each } \mathbf{x} \in X\}$. We have $X^{\perp} \in \mathcal{L}$, $\left(X^{\perp}\right)^{\perp} = X$, $X \cap X^{\perp} = \{\mathbf{0}\}$, $X + X^{\perp} = \mathbb{R}^n$. Let $\mathbf{z} \in \mathbb{R}^n$ and $X \in \mathcal{L}$, then there is a unique $\mathbf{x} \in X$ such that $\|\mathbf{z} - \mathbf{x}\| \leq \|\mathbf{z} - \mathbf{y}\|$ for all $\mathbf{y} \in X$. $\mathbf{x}$ is called *the orthogonal projection of* $\mathbf{z}$ *onto* X. Let $\delta_{ii} = 1$ but $\delta_{ij} = 0$ for $i \neq j$. $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_m$ form an *orthonormal basis* for $X \in \mathcal{L}$ if $\mathbf{s}_i\dagger\mathbf{s}_j = \delta_{ij}$ and if any vector $\mathbf{x} \in X$ can be represented as a linear combination of the basis vectors: $\mathbf{x} = \sum_{i=1}^{m} x_i \mathbf{s}_i$, where the numbers $x_1, x_2, ..., x_m$ are uniquely determined by $\mathbf{x}$ and $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_m$. The choice of an orthonormal basis for a vector space is arbitrary. However, the inner product of two vectors is independent of the orthonormal basis chosen. We shall refer to a normalized vector, $\mathbf{s} \in \mathbb{R}^n$, as a *state*

[84]More generally, in $\mathbb{C}^n$, $\mathbf{x}\dagger$ is the *adjoint*, of $\mathbf{x}$. For example, in $\mathbb{C}^2$, if $\mathbf{x} = \begin{bmatrix} r_1 e^{i\theta_1} \\ r_2 e^{i\theta_2} \end{bmatrix}$, where r_1, θ_1, r_2, θ_2 are real and $i = \sqrt{-1}$, then $\mathbf{x}\dagger = \begin{bmatrix} r_1 e^{-i\theta_1} & r_2 e^{-i\theta_2} \end{bmatrix}$.

[85]More generally, in $\mathbb{C}^n$, $\mathbf{x}\dagger\mathbf{y} = \sum_{i=1}^{n} x_i^* y_i$, where, if $x = re^{i\theta}$, $r, \theta \in \mathbb{R}$, then $x^* = re^{-i\theta}$.

[86]In *Dirac notation*, $\mathbf{x} = |x\rangle$, $\mathbf{x}\dagger = \langle x|$, $\mathbf{x}\dagger\mathbf{y} = \langle x|y\rangle$, $\|\mathbf{x}\| = \sqrt{\mathbf{x}\dagger\mathbf{x}} = \sqrt{\langle x|x\rangle}$.

vector. In particular, if $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n$ form an orthonormal basis for $\mathbb{R}^n$, then we shall refer to these as *eigenstates*. Note that if $\mathbf{s} = \sum_{i=1}^{n} s_i \mathbf{s}_i$, then $\mathbf{s}$ is a state vector if, and only if, $\|\mathbf{s}\| = 1$, equivalently, if, and only if, $\mathbf{s}\dagger\mathbf{s} = \sum_{i=1}^{n} s_i s_i = 1$. Let $X \in \mathcal{L}$. Let $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_m$ form an orthonormal basis for X. Extend $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_m$ to an orthonormal basis, $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_m, ..., \mathbf{s}_n$, for $\mathbb{R}^n$ (this can always be done). Then $\mathbf{s}_{m+1}, ..., \mathbf{s}_n$ form an orthonormal basis for the the orthogonal complement, $X^\perp$, of X. Let $\mathbf{z} = \sum_{i=1}^{n} z_i \mathbf{s}_i \in \mathbb{R}^n$. Then $\sum_{i=1}^{m} z_i \mathbf{s}_i$ is the orthogonal projection of $\mathbf{z}$ onto X and $\sum_{i=m1}^{n} z_i \mathbf{s}_i$ is the orthogonal projection of $\mathbf{z}$ onto $X^\perp$.

We will represent the state of the known Ellsberg urn (K) by a normalized vector in $\mathbb{R}^2$ and the unknown Ellsberg urn (U) by a normalized vector in $\mathbb{R}^4$.

4.4.2 State of a system, events and quantum probability measures

The state of a system (physical, biological or social) is represented by a normalized vector, $\mathbf{s} \in \mathbb{R}^n$, i.e., $\|\mathbf{s}\| = 1$. The set of events is the set, $\mathcal{L}$, of vector subspaces of $\mathbb{R}^n$. $\{\mathbf{0}\}$ is the impossible event and $\mathbb{R}^n$ is the certain event. $X^\perp \in \mathcal{L}$ is the complement of the event $X \in \mathcal{L}$. If $X, Y \in \mathcal{L}$ then $X \cap Y$ is the conjunction of the events X and Y; $X + Y$ is the event where either X occurs or Y occurs or both (if $X, Y \in \mathcal{L}$ then, in general, $X \cup Y \notin \mathcal{L}$). Recall that in a σ-algebra of subset of a set, the distributive law: $X \cap (YUZ) = (X \cap Y) \cup (X \cap Z)$, and its dual[87], holds. However, its analogue for $\mathcal{L}$: $X \cap (Y + Z) = (X \cap Y) + (X \cap Z)$, and its dual[88], fails to hold in general. Consequently, the law of total probability also fails to hold in general. The failure of the distributive laws to hold in $\mathcal{L}$ has profound consequences. This non-distributive nature of $\mathcal{L}$ is the key to explaining many paradoxes of human behaviour. $F : \mathcal{L} \to [0, 1]$ is *additive* if $F\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} F(X_i)$, where $X_i \in \mathcal{L}$ and $X_i \cap X_j = \{\mathbf{0}\}$ for $i \neq j$. A *quantum probability measure* is an additive measure, $P : \mathcal{L} \to [0, 1]$, $P(\{\mathbf{0}\}) = 0$, $P(\mathbb{R}^n) = 1$. If a number can be interpreted as either a classical probability or a quantum probability, then we shall simply refer to it as a probability. Otherwise, we shall refer to it as either a classical probability or a quantum probability, whichever is the case.

4.4.3 Random variables and expected values

Let $\mathcal{L}$ be the set of all vector subspaces of $\mathbb{R}^n$. A *random quantum variable* is a mapping, $f : \mathbb{R}^n \to \mathbb{R}$ satisfying: $\{\varphi \in \mathbb{R}^n : f(\varphi) \leq r\} \in \mathcal{L}$ for each $r \in \mathbb{R}$.

[87] $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$

[88] $X + (Y \cap Z) = (X + Y) \cap (X + Z)$

A random quantum variable, f, is *non-negative* if $f(\varphi) \geq 0$ for each $\varphi \in \mathbb{R}^n$. For two random quantum variables, f, g, we write $f \leq g$ if $f(\varphi) \leq g(\varphi)$ for each $\varphi \in H$. A random quantum variable, f, is *simple* if its range is finite. For any random quantum variable, f, and any $\varphi \in \mathbb{R}^n$, let $f^+(\varphi) = \max\{0, f(\varphi)\}$ and $f^-(\varphi) = -\min\{0, f(\varphi)\}$. Then, clearly, f^+ and f^- are both non-negative random quantum variables and $f(\varphi) = f^+(\varphi) - f^-(\varphi)$, for each $\varphi \in \mathbb{R}^n$. We write this as $f = f^+ - f^-$.

Let f be a simple random quantum variable with range $\{f_1, f_2, ..., f_n\}$. Let $X_i = \{\varphi \in \mathbb{R}^n : f(\varphi) = f_i\}$. Then $X_i \in \mathcal{L}$, $X_i \cap X_j = \{0\}$ for $i \neq j$ and $\sum_{i=1}^n X_i = \mathbb{R}^n$. Then the *expected value* of the simple random quantum variable, f, is $E(f) = \sum_{i=1}^n f_i P(X_i)$. The *expected value* of the non-negative random quantum variable, g, is

$$E(g) = \sup\{E(f) : f \leq g \text{ is a simple random quantum variable}\}.$$

Note that $E(g)$ may be infinite. If $f = f^+ - f^-$ is an arbitrary random quantum variable such that not both $E(f^+)$ and $E(f^-)$ are infinite, then the expected value of f is $E(f) = E(f^+) - E(f^-)$. Note that $E(f)$ can be $-\infty$, finite or ∞. However, if both $E(f^+)$ and $E(f^-)$ are both infinite then $E(f)$ is undefined (because $\infty - \infty$ is undefined).

4.4.4 Transition amplitudes and probabilities

Suppose $\varphi, \chi \in \mathbb{R}^n$ are two states (thus, they are normalized: $\|\varphi\| = \|\chi\| = 1$). $\varphi \to \chi$ symbolizes the transition from φ to χ. Then, by definition, the *amplitude* of $\varphi \to \chi$ is given by $A(\varphi \to \chi) = \varphi\dagger\chi$. Its quantum probability is $P(\varphi \to \chi) = (\varphi\dagger\chi)^2$.[89]

Consider the state $\varphi \in \mathbb{R}^n$ ($\|\varphi\|=1$). The occurrence of the event $X \in \mathcal{L}$ causes a transition, $\varphi \to \psi$. The new state, ψ ($\|\psi\|=1$), can be found as follows. Let π be the orthogonal projection of φ onto X (recall subsection 4.4.1). Suppose that $\pi \neq \mathbf{0}$ (if $\pi = \mathbf{0}$, then π and X are incompatible, that is, if X occurs then the transition $\varphi \to \psi$ is impossible). Then $\psi = \frac{\pi}{\|\pi\|}$ is the new *state conditional on* X.

4.4.5 Born's rule

We can now give the empirical interpretation of the state vector. Consider a physical, biological or social system. On measuring a certain observable pertaining to the system, this observable can take the value $v_i \in \mathbb{R}$ with probability $p_i \geq 0$, $\sum_{i=1}^n p_i =$

[89] In $\mathbb{C}^n$, $P(\varphi \to \chi) = (\varphi\dagger\chi)(\varphi\dagger\chi)^*$. However, as we are working in $\mathbb{R}^n$, $(\varphi\dagger\chi) = (\varphi\dagger\chi)^*$.

1. To model this situation, let $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n$ form an orthonormal bases for $\mathbb{R}^n$. Take $\mathbf{s}_i$ to be the state (eigenstate) where the observable takes the value (eigenvalue) v_i for sure. Consider the general state $\mathbf{s} = \sum_{i=1}^{n} s_i \mathbf{s}_i$. If the act of measurement gives the value v_i for the observable, then this implies that the act of measurement has caused the transition $\mathbf{s} \to \mathbf{s}_i$. The probability of the transition $\mathbf{s} \to \mathbf{s}_i$ is $P(\mathbf{s} \to \mathbf{s}_i) = (\mathbf{s} \dagger \mathbf{s}_i)^2 = s_i^2 = p_i$. Thus, in the representation of the state of the system by $\mathbf{s} = \sum_{i=1}^{n} s_i \mathbf{s}_i$, s_i^2 is the probability of obtaining the value v_i on measurement.[90]

4.4.6 Feynman's first rule (single path)

Let φ, χ, ψ be three states. $\varphi \to \chi \to \psi$ symbolizes the transition from φ to χ followed by the transition from χ to ψ. The amplitude of $\varphi \to \chi \to \psi$ is then the *product*, $A(\varphi \to \chi \to \psi) = A(\varphi \to \chi) A(\chi \to \psi) = (\varphi \dagger \chi)(\chi \dagger \psi)$, of the amplitudes of $\varphi \to \chi$ and $\chi \to \psi$. The quantum probability of the transition, $\varphi \to \chi \to \psi$, is then $P(\varphi \to \chi \to \psi) = (A(\varphi \to \chi \to \psi))^2 = ((\varphi \dagger \chi)(\chi \dagger \psi))^2 = (\varphi \dagger \chi)^2 (\chi \dagger \psi)^2$, i.e., the product of the respective probabilities. This can be extended to any number of multiple transitions along a single path.

4.4.7 Feynman's second rule (multiple indistinguishable paths)

Suppose that the transition from φ to ψ can follow any of two paths:
$\varphi \to \chi_1 \to \psi$ or $\varphi \to \chi_2 \to \psi$. Furthermore, and this is crucial, assume that which path was followed is *not* observable. First, we calculate the amplitude of $\varphi \to \chi_1 \to \psi$, using Feynman's first rule. We also calculate the amplitude of $\varphi \to \chi_2 \to \psi$, using, again, Feynman's first rule. To find the amplitude of $\varphi \to \psi$ (via χ_1 or χ_2) we *add the two amplitudes*. The amplitude of $\varphi \to \psi$ is then $(\varphi \dagger \chi_1)(\chi_1 \dagger \psi) + (\varphi \dagger \chi_2)(\chi_2 \dagger \psi)$. Finally, the probability of the transition $\varphi \to \psi$ (via χ_1 or χ_2) is $((\varphi \dagger \chi_1)(\chi_1 \dagger \psi) + (\varphi \dagger \chi_2)(\chi_2 \dagger \psi))^2 = (\varphi \dagger \chi_1)^2 (\chi_1 \dagger \psi)^2 + (\varphi \dagger \chi_2)^2 (\chi_2 \dagger \psi)^2 + 2((\varphi \dagger \chi_1)(\chi_1 \dagger \psi)(\varphi \dagger \chi_2)(\chi_2 \dagger \psi))$.

4.4.8 Feynman's third rule (multiple distinguishable paths)

Suppose that the transition from φ to ψ can follow any of two paths:
$\varphi \to \chi_1 \to \psi$ or $\varphi \to \chi_2 \to \psi$. Furthermore, and this is crucial, assume that which path was followed *is* observable (although it might not actually be observed). First, we calculate the quantum probability of $\varphi \to \chi_1 \to \psi$, using Feynman's first rule. We also calculate the quantum probability of $\varphi \to \chi_2 \to \psi$, using, again,

[90]More generally, if we use $\mathbb{C}^n$, then $s_i^* s_i$, is the probability of obtaining the value v_i on measurement.

Feynman's first rule. To find the total quantum probability of $\varphi \to \psi$ (via χ_1 or χ_2) we *add the two probabilities*. The quantum probability of $\varphi \to \psi$ is then $(\varphi\dagger\chi_1)^2 (\chi_1\dagger\psi)^2 + (\varphi\dagger\chi_2)^2 (\chi_2\dagger\psi)^2$.

Comparing the last expression with its analogue for Feynman's second rule, we see the absence here of the term $2\left((\varphi\dagger\chi_1)(\chi_1\dagger\psi)(\varphi\dagger\chi_2)(\chi_2\dagger\psi)\right)$. This is called the *interference term*. Its presence or absence has profound implications in both quantum physics and quantum decision theory.

The Feynman rules play a role in quantum probability theory analogous to the rule played by Bayes' law and the law of total probability in classical theory.

4.5 A quantum decision model of the Ellsberg paradox

4.5.1 Quantum decision model

Urn U contains two balls labeled 1 if it is in state $\mathbf{s}_1$ (recall section 4.3). It contains one ball labeled 1 and the other labeled 2 if it is either in state $\mathbf{s}_2$ or in state $\mathbf{s}_3$. In state $\mathbf{s}_4$ both balls are labeled 2. We represent these states in $\mathbb{R}^4$ by the orthonormal basis:

$$\mathbf{s}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{s}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{s}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{s}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Let $\mathbf{s}$ give the initial state of urn U (unknown composition). Then *Born's rule* leads to:

$$\mathbf{s} = p\mathbf{s}_1 + \sqrt{p(1-p)}\mathbf{s}_2 + \sqrt{(1-p)p}\mathbf{s}_3 + (1-p)\mathbf{s}_4, \tag{4.3}$$

where there is a probability p^2 that ball 1 is drawn in each round (state $\mathbf{s}_1$), a probability $p(1-p)$ that ball 1 is drawn in round 1 then ball 2 is drawn in round 2 (state $\mathbf{s}_2$), a probability $(1-p)p$ that ball 2 is drawn in round 1 then ball 1 is drawn in round 2 (state $\mathbf{s}_3$) and, finally, a probability $(1-p)^2$ that ball 2 is drawn in each round (state $\mathbf{s}_4$).

Let the event that ball 1 is drawn from urn U be denoted by $\mathbf{t}$. We now calculate the probability of event $\mathbf{t}$.

Proposition 4.2. *(al-Nowaihi and Dhami, 2017): If the probability of drawing ball* 1 *from the known urn K is p, then the quantum probability of drawing ball* 1 *from the unknown urn U is*

$$Q(p) = \frac{5p^3 - 8p^2 + 4p}{2-p}. \tag{4.4}$$

Proof of Proposition 4.2: The role played by the *law of reciprocity* was only implicit in al-Nowaihi & Dhami (2017). Here we make it explicit. In general, the law of total probability, recall (4.1) above, is not valid in quantum probability theory. See Busemeyer & Bruza (2012), chapter 1, pp. 5. Instead, we use the *Feynman's rules* (see Busemeyer & Bruza (2012), chapter 1, pp. 13) and the *law of reciprocity* (see Busemeyer & Bruza (2012), chapter 2, pp. 39). In our case, working in the Hilbert space $\mathbb{C}^4$ gives the same results as working in $\mathbb{R}^4$, as can be verified by direct calculation. Hence, for simplicity, we shall work in the Hilbert space $\mathbb{R}^4$. Recall that the state of a quantum system is given by normalized vector, $\mathbf{s}$, in Hilbert space, i.e., $\mathbf{s}\dagger\mathbf{s} = (\mathbf{s}\dagger)\,\mathbf{s} = 1$, where $\mathbf{s}\dagger$ is the *conjugate transpose* of $\mathbf{s}$ (in our case, simply the transpose of $\mathbf{s}$, since we are working in $\mathbb{R}^4$). We give the proof in several stages.

Reciprocity Let $\mathbf{t}$ be the state where ball 1 is drawn from U. We wish to calculate the probability, $P(\mathbf{s} \to \mathbf{t})$, of the transition $\mathbf{s} \to \mathbf{t}$. By the quantum law of reciprocity, $P(\mathbf{s} \to \mathbf{t}) = P(\mathbf{t} \to \mathbf{s})$, both being equal to $(\mathbf{s}\dagger\mathbf{t})^2$. Recall we are working in a real Hilbert space. For a complex Hilbert space, we would have $P(\mathbf{s} \to \mathbf{t}) = P(\mathbf{t} \to \mathbf{s}) = (\mathbf{s}\dagger\mathbf{t})(\mathbf{s}\dagger\mathbf{t})^*$, where $(\mathbf{s}\dagger\mathbf{t})^*$ is the complex conjugate of $\mathbf{s}\dagger\mathbf{t}$. But $P(\mathbf{t} \to \mathbf{s})$ is the probability of the state of U conditional on drawing ball 1 from U. Let $\mathbf{w}$ be this state. To find $\mathbf{w}$, we first project $\mathbf{s}$ onto the subspace spanned by $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3\}$, then normalize. This gives

$$\mathbf{w} = \sqrt{\frac{p}{2-p}}\mathbf{s}_1 + \sqrt{\frac{1-p}{2-p}}\mathbf{s}_2 + \sqrt{\frac{1-p}{2-p}}\mathbf{s}_3. \tag{4.5}$$

Feynman's rules To arrive at the state, $\mathbf{w}$, the state of urn U conditional on ball 1 being drawn, we must follow one of the three paths:

1. $\mathbf{s} \to \mathbf{s}_1 \to \mathbf{w}$,

2. $\mathbf{s} \to \mathbf{s}_2 \to \mathbf{w}$.

3. $\mathbf{s} \to \mathbf{s}_3 \to \mathbf{w}$.

Using Feynman's first rule (single path), $A(\mathbf{s} \to \mathbf{s}_i \to \mathbf{w})$
$= A(\mathbf{s} \to \mathbf{s}_i)\, A(\mathbf{s}_i \to \mathbf{w})$, the relevant transition amplitudes are:
$A(\mathbf{s} \to \mathbf{s}_1) = \mathbf{s}\dagger\mathbf{s}_1 = p$, $A(\mathbf{s}_1 \to \mathbf{w}) = \mathbf{s}_1\dagger\mathbf{w} = \sqrt{\frac{p}{2-p}}$, $A(\mathbf{s} \to \mathbf{s}_1 \to \mathbf{w}) = \sqrt{\frac{p^3}{2-p}}$.
$A(\mathbf{s} \to \mathbf{s}_2) = \mathbf{s}\dagger\mathbf{s}_2 = \sqrt{p(1-p)}$, $A(\mathbf{s}_2 \to \mathbf{w}) = \mathbf{s}_2\dagger\mathbf{w} = \sqrt{\frac{1-p}{2-p}}$, $A(\mathbf{s} \to \mathbf{s}_2 \to \mathbf{w})$
$= (1-p)\sqrt{\frac{p}{2-p}}$.
$A(\mathbf{s} \to \mathbf{s}_3) = \mathbf{s}\dagger\mathbf{s}_3 = \sqrt{p(1-p)}$, $A(\mathbf{s}_3 \to \mathbf{w}) = \mathbf{s}_3\dagger\mathbf{w} = \sqrt{\frac{1-p}{2-p}}$, $A(\mathbf{s} \to \mathbf{s}_3 \to \mathbf{w})$
$= (1-p)\sqrt{\frac{p}{2-p}}$.

We shall treat the paths $\mathbf{s} \to \mathbf{s}_2 \to \mathbf{w}$ and $\mathbf{s} \to \mathbf{s}_3 \to \mathbf{w}$ as indistinguishable from each other but both distinguishable from path $\mathbf{s} \to \mathbf{s}_1 \to \mathbf{w}$. Our argument for this is as follows. The path $\mathbf{s} \to \mathbf{s}_1 \to \mathbf{w}$ results in urn U containing two balls labeled 1. This is clearly distinguishable from paths $\mathbf{s} \to \mathbf{s}_2 \to \mathbf{w}$ and $\mathbf{s} \to \mathbf{s}_3 \to \mathbf{w}$, each of which result in urn U containing one ball labeled 1 and one ball labeled 2. From examining urn U, it is impossible to determine whether this arose by selecting ball 1 first (path $\mathbf{s} \to \mathbf{s}_2 \to \mathbf{w}$), then ball 2 (path $\mathbf{s} \to \mathbf{s}_3 \to \mathbf{w}$), or the other way round.

We apply Feynman's second rule (multiple indistinguishable paths) to find the amplitude of the transition $\mathbf{s} \to \mathbf{w}$, via $\mathbf{s}_2$ or via $\mathbf{s}_3$. We add the amplitudes of these two paths. Thus, $A(\mathbf{s} \to \mathbf{w})$, via $\mathbf{s}_2$ or $\mathbf{s}_3$ is $A(\mathbf{s} \to \mathbf{s}_2 \to \mathbf{w}) + A(\mathbf{s} \to \mathbf{s}_3 \to \mathbf{w}) = 2(1-p)\sqrt{\frac{p}{2-p}}$. The probability of this transition is $\left(2(1-p)\sqrt{\frac{p}{2-p}}\right)^2 = \frac{4p(1-p)^2}{2-p}$. The probability of the transition $\mathbf{s} \to \mathbf{s}_1 \to \mathbf{w}$ is $\left(\sqrt{\frac{p^3}{2-p}}\right)^2 = \frac{p^3}{2-p}$. We apply Feynman's third rule (multiple distinguishable paths) to get the total probability of the transition $\mathbf{s} \to \mathbf{w}$, via all paths. We add the two probabilities. This gives $P(\mathbf{s} \to \mathbf{w}) = \frac{p^3}{2-p} + \frac{4p(1-p)^2}{2-p} = \frac{5p^3-8p^2+4p}{2-p}$.

Quantum probability Recall that $\mathbf{s}$ is the initial state of urn U, $\mathbf{t}$ is the state in which ball 1 is drawn and $\mathbf{w}$ is the state of urn U conditional on ball 1 having been drawn. We wish to calculate the probability, $P(\mathbf{s} \to \mathbf{t})$, of the transition $\mathbf{s} \to \mathbf{t}$. By the quantum law of reciprocity, $P(\mathbf{s} \to \mathbf{t}) = P(\mathbf{t} \to \mathbf{s})$. But $P(\mathbf{t} \to \mathbf{s})$ is the probability of the state of U conditional on drawing ball 1 from U. We have already calculated this to be $\frac{5p^3-8p^2+4p}{2-p}$.

Thus, if the probability of drawing ball 1 from the known urn K is p, then the quantum probability of drawing ball 1 from the unknown urn U is

$$Q(p) = \frac{5p^3 - 8p^2 + 4p}{2-p}.$$

This completes the proof of Proposition 4.2. ■

Suppose the contents of the unknown urn U are kept fixed but a new known urn, K_i, is constructed so that the probability of drawing ball 1 from urn K_i is now $Q(p)$. In section 4.5.2, below, we shall prove that subject i should be indifferent between U and K_i, i.e., $Q(p)$ is the *matching probability* of p.

From (4.4), we get

$$Q(0.1) = 0.17105,\ Q(0.5) = 0.41667,\ Q(0.9) = 0.69545, \tag{4.6}$$

in close agreement with the evidence given by Dimmock et al. (2015) and our own evidence given later in this paper.

The following results are easily established.

$$Q(0) = 0,\ Q(1) = 1.$$

$$Q(p) + Q(1-p) < 1 \text{ for all } 0 < p < 1.$$

$$\underset{p\to 0}{lim} Q(p) = 0,\ \underset{p\to 0}{lim}\frac{Q(p)}{p} = 2,\ \underset{p\to 1}{lim}\frac{Q(p)}{p} = 1.$$

$$p < 0.4 \Rightarrow Q(p) > p,\ p = 0.4 \Rightarrow Q(p) = p,\ p > 0.4 \Rightarrow Q(p) < p. \tag{4.7}$$

Note that (4.4) is parameter free. By contrast, the probably most successful approach to ambiguity, *source dependent theory* (Abdellaoui et al., 2011; Kothiyal et al., 2014; Dimmock et al., 2015), requires the specification of two probability weighting functions, $w_K(p)$ and $w_U(p)$, one for urn K and one for urn U. These require the estimation of at least two parameters. For example, using Prelec (1998) probability weighting functions, $w_K(p) = e^{-\beta_K(-\ln p)^{\alpha_K}}$ and $w_U(p) = e^{-\beta_U(-\ln p)^{\alpha_U}}$, requires estimating two parameters: $\alpha = \frac{\ln\beta_U - \ln\beta_K}{\alpha_K}$, $\beta = \frac{\alpha_U}{\alpha_K}$.[91]

4.5.2 Quantum probabilities are matching probabilities

If p is the probability of drawing ball 1 from the known urn K, then $Q(p)$, given by (4.4), is the quantum probability of drawing ball 1 from the unknown urn U. Let u_i be the utility function of a subject, i, participating in the Ellsberg experiment as perceived by the subject (recall section 4.3). Normalize u_i so that $u_i(0) = 0$. The subject wins the sum of money, $v > 0$, if ball 1 is drawn from the unknown urn U, but zero if ball 2 is drawn from that same urn. Hence, her projective expected utility (in the sense of La Mura, 2009) is

$$Q(p)\,u_i(v)\,. \tag{4.8}$$

Now construct a new known urn K_i from which ball 1 is drawn with probability $Q(p)$. Her projective expected utility is

$$Q(p)\,u_i(v)\,. \tag{4.9}$$

[91]Let $u_i(v)$ be the utility of subject i, normalized so that $u_i(0) = 0$. From the definition of matching probabilities, we have $w_K(m(p))\,u_i(v) = w_U(p)\,u_i(v)$. This gives $\ln(-\ln m(p)) = \alpha + \beta\ln(-\ln p)$.

Hence, from (4.8) and (4.9), $Q(p)$ is the matching probability for p. Thus, subject i is ambiguity *averse*, *neutral* or *seeking* according to $Q(p)$ being *less than*, *equal to* or *greater than* p.

From (4.7), it then follows that:

$$\begin{aligned} p < 0.4 &\Rightarrow Q(p) > p: \text{ ambiguity seeking,} \\ p = 0.4 &\Rightarrow Q(p) = p: \text{ ambiguity neutral,} \\ p > 0.4 &\Rightarrow Q(p) < p: \text{ ambiguity averse.} \end{aligned}$$

Thus, our model is in agreement with stylized fact 1 (*insensitivity*).

4.6 Experimental design

Our subjects were 295 undergraduate students from Qingdao Agricultural University in China. They attended 8 sessions; no one participated in more than one session. The experimental instructions are given in the Appendix C.1.

Our treatment was a paper-based classroom experiment. There were three tasks, *Task 1*, *Task 2* and *Task 3*, that were, respectively, designed to implement the three cases $p = 0.5$, $p = 0.1$, $p = 0.9$ (see (4.6)). Each task required two tables to be completed. The materials for each task were handed out at the beginning of that task and collected before the next task started.

In each task, there is one known urn (Box K) and one unknown urn (Box U). The composition of the 100 colored balls of k different colors in Box K is known; varying this composition gives us the three cases $p = 0.5$, 0.1, 0.9. Box U contains 100 colored balls of the same colors as in Box K, but in unknown proportions. The composition of Box U is randomly decided at the end of the experiment in the following way. Each ball is equally likely to be drawn. The random draw follows the uniform distribution. For example, in Task 2, there are in total 10 different colors. A priori, each color is equally likely to be drawn. Thus, at each stage of the construction of Box U, each color has a probability 0.1 of being the color of the next ball to be placed in Box U. There can be from 0 to 100 balls of any particular color but subject to the restriction that the total number of balls in Box U is 100 balls. The prize for drawing a winning-color ball is 10 *Yuan* whether it is drawn from Box K or Box U. We now explain the three tasks.

1. In Task 1, there are 50 purple balls and 50 yellow balls in Box K, and *purple* is the winning color ($p = 0.5$).[92] The decision maker is shown two tables. In

[92]These are the same colors as chosen by Dimmock et al. (2015).

Table 1, the choices are to express a preference to receive a monetary amount x for sure, express indifference between x or betting that a purple ball will be drawn from Box K, or express a preference for betting that a purple ball will be drawn from Box K. The monetary amount is varied from $x = 0$ to $x = 10$ and subjects have to state a choice in each case.[93] Box U has 100 balls that are either purple or yellow but the proportions are unknown; as explained above. Table 2 replaces Box K in Table 1 with Box U but it is otherwise identical. At the end of the experiment, one of the choices from Task 1 is picked at random to be played for real.

2. In Task 2, there are 10 different colors (including purple) in Box K, and *purple* is the winning color ($p = 0.1$). Box U has 100 balls of the same 10 colors but in unknown proportions. The remaining procedure is as described in Task 1.

3. In Task 3, there are 10 different colors (including purple) in Box K, and the winning color is any ball that is not purple ($p = 0.9$). Box U has 100 balls of the same 10 colors but in unknown proportions. The remaining procedure is as described in Task 1.

4.7 Experimental results

Consider a sample of N subjects. Choose a probability, p, for drawing a winning ball from urn K. For each of these N subjects, we elicit their matching probability. Find the matching probability, $m_i(p)$, for each subject, i, $i = 1, 2, ..., N$, the sample average, $m(p) = \frac{1}{N}\sum_{i=1}^{N} m_i(p)$ and the sample variance, $s^2 = \frac{1}{N}\sum_{i=1}^{N} (m_i(p) - m(p))^2$. The *t-statistic* is $t = \frac{m(p) - Q(p)}{s/\sqrt{N}}$, where $Q(p)$ is the quantum prediction.

It might not be surprising to see much unsystematic variability in the matching probabilities, $m_i(p)$, across the sample.[94] However, if our quantum model is correct, then, for large N, this unsystematic variability should be largely cancelled out in aggregate. Hence, we would expect t to be approximately normally distributed with mean 0 and variance 1.[95] For ease of reference, we give the critical values for each of the conventional significance levels (10%, 5%, 1%) for a two-tailed test for the standard normal distribution in Table 4.1, below.

[93] The experiments were conducted in China, so the monetary amount is in units of Chinese Yuan.

[94] Sufficient conditions for this are that $m_i(p) = m(p) + \epsilon_i$, where $E(\epsilon_i) = 0$, $i = 1, 2, ..., N$ and ϵ_i and ϵ_j are identically and independently for $i \neq j$.

[95] See, for example, Chapter 5 of Wooldridge (2015).

Table 4.1: Significance levels and the corresponding critical values.

Significance level	Critical value
10%	± 1.64
5%	± 1.96
1%	± 2.58

We collected in total 19470 ($= 11 \times 2 \times 3 \times 295$) data points.[96] There were 259, 262 and 263 consistent decision makers for the $p = 0.1$, $p = 0.5$ and $p = 0.9$ cases, respectively[97].

We chose these particular probabilities for two reasons: (1) to facilitate comparison with Dimmock et al. (2015). (2) insensitivity becomes more marked the further we move away form $p = 0.5$. Hence, testing for p near the end points, 0 and 1, gives a more stringent test of the theory than testing for p in the middle range.

We estimated the cash equivalents for the decisions in the two tables in Appendix C.1 in the following way. If there is one unique tick in the "Indifference" column in the table, then the cash equivalent is the corresponding amount of money s/he receives for sure (x); On the other hand, if there is no tick in the "Indifference" column, then the cash equivalent is estimated by the midpoint between the lowest amount of money that is preferred to the uncertain bet, and the highest amount of money for which the bet was preferred; we are following the methodology in study 2 of Fox and Tversky (1995).

To find the matching probability with the cash equivalents that we obtained, it is necessary to assume a form for the utility function.[98] We use the power function[99] for the utility of player i,

$$u_i(x) = x^{\sigma_i}, x \geq 0, \sigma_i > 0. \tag{4.10}$$

Let v be the monetary payment to a subject if a winning ball is drawn. Let p be the probability of selecting a winning ball from the known urn (K). Let $m_i(p)$ be the matching probability, for subject i, of selecting a winning ball from the

[96]This indicates 11 data points (for the 11 rows of Tables 1 and 2; see Appendix C.1); 2 Tables corresponding to the known and unknown urns (Tables 1 and 2 in Appendix C.1); 3 tasks (Task 1, Task 2, and Task 3); and 295 subjects in the experiment.

[97]We discarded the inconsistent decision makers from the analysis as follows. We discarded data with the following two patterns: firstly, choosing more than once in the "Indifference" column in the table; secondly, choosing back and forth in any two or three columns. For the first case, we cannot identify the unique cash equivalent; while, it seems that the subjects with the second behavioral pattern don't show a clear ambiguity attitude. This left us with over 250 subjects.

[98]This is necessary in the methodology in study 2 of Fox and Tversky (1995); which is incentive compatible. It is not necessary in the methodology of Dimmock et al. (2015). However, the latter is not incentive compatible.

[99]Recall subsection 4.1.4 of the Introduction.

unknown urn (U). Additionally, the monetary valuation of the known urn (K) to subject i is denoted by v_{iK}, while the monetary valuation of the unknown urn (U) to subject i is denoted by v_{iU}. v_{iK} and v_{iU} are respectively the cash equivalents in the corresponding tables (recall the cash equivalents explained above).

Firstly, for the known urn (K), we have

$$(v_{iK})^{\sigma_i} = p(v)^{\sigma_i}. \tag{4.11}$$

Solve it for σ_i, to get

$$\sigma_i = \frac{-\ln p}{\ln v - \ln v_{iK}}, \tag{4.12}$$

where all quantities on the right hand side are known. Therefore, σ_i can be calculated using known quantities. Specifically, $v = 10$ *Yuan*; $p = 0.1$, $p = 0.5$ or $p = 0.9$ in the three cases; v_{iK} is the cash equivalent that we determine from the experiment.[100]

Similarly, for the unknown urn (U), we have

$$(v_{iU})^{\sigma_i} = m_i\,(p)\,(v)^{\sigma_i}. \tag{4.13}$$

Solve for $m_i\,(p)$, to get

$$m_i\,(p) = (\frac{v_{iU}}{v})^{\sigma_i}. \tag{4.14}$$

Substitute from (4.12) into (4.14) to get

$$m_i\,(p) = (\frac{v_{iU}}{v})^{\frac{-\ln p}{\ln v - \ln v_{iK}}}. \tag{4.15}$$

Since all quantities on the right hand side of (4.15) are known, the matching probability can be found (recall v_{iU} is the cash equivalent). Following this approach, we find the mean matching probabilities, $m\,(p)$, and standard deviations, which are listed in Table 4.2, below. The fifth column of Table 4.2 shows the theoretical predictions for the three matching probabilities, $Q(0.1)$, $Q(0.5)$, and $Q(0.9)$, respectively.[101]

Table 4.2, below, shows that the theoretically predicted matching probabilities are quite close to the mean values we obtained from our experiments.

Our null and alternative hypotheses are: $H_0 : m\,(p) = Q\,(p)$ and $H_1 : m\,(p) \neq Q\,(p)$. From Tables 4.1 and 4.2, 1.4758, 1.4437, 2.3906 are all less than 2.58. Thus,

[100] One subject chose $v_{iK} = v = 10$, for $p = 0.9$. Since the denominator in (4.12) would then be zero for these values, we discarded this observation.

[101] The theoretically predicted values are found by substituting the values of p, 0.1, 0.5, and 0.9, respectively, into (4.4).

Table 4.2: t-test for the means.

Matching probability	Mean	Standard deviation	Sample size	Quantum probability Q	t-stat
$m\,(0.1)$	0.1864	0.1708	259	0.1711	1.4437
$m\,(0.5)$	0.4038	0.1416	262	0.4167	-1.4758
$m\,(0.9)$	0.7258	0.2056	263	0.6955	2.3906

our experimental results fail to reject our quantum model at the 1% level of significance. Since $m\,(0.1) > 0.1$, $m\,(0.5) < 0.5$, $m\,(0.9) < 0.9$, we find ambiguity seeking for the low probability but ambiguity aversion for the medium and high probabilities.

4.8 Demographic results

In their answers to question 8 on the post-experimental questionnaire (Appendix C.2), only 4 out of the 295 subjects reported that color affected their decisions. In their answers to question 6, almost none reported prior experience with similar experiments in the past. In their answers to question 4, Degree of study, all students simply gave “undergraduate”, thus giving us no useful information. From the answers to question 3 (Field of study), we obtained the data for economics/non-economics. Not surprisingly, we found high colinearity between year of study and age, so we have not reported the latter.

4.8.1 Mann-Whitney U tests

We used two-sided Mann-Whitney U test (nonparametric test) to examine if the demographic characteristics in Appendix C.2 affected the subjects’ reported matching probabilities for $p = 0.1$, $p = 0.5$ and $p = 0.9$ in our treatment. The results are shown in Table 4.3. At the 1% level, no significant differences were found between any of the two groups (male/female; economics/non-economics students; statistics/non-statistics students).

4.8.2 *t*-tests

For each demographic group, we also performed a t-test to see if the average reported matching probability, $m\,(p)$, differed significantly from the predicted value of the quantum probability, $Q\,(p)$. We report the results in Table 4.4, below. The only group that showed a significant difference was the group of students with prior

Table 4.3: Mann-Whitney U test results.

Group	Matching probability	MWU p-value	Sig diff
	$m\,(0.1)$	0.9533	No
Male vs. Female	$m\,(0.5)$	0.2825	No
	$m\,(0.9)$	0.5205	No
	$m\,(0.1)$	0.8941	No
Econ vs. Non-econ	$m\,(0.5)$	0.7529	No
	$m\,(0.9)$	0.1230	No
	$m\,(0.1)$	0.0496	No*
Stats vs. Non-stats	$m\,(0.5)$	0.2053	No
	$m\,(0.9)$	0.7413	No
	$m\,(0.1)$	0.2944	No
Year 1 vs. Year 2	$m\,(0.5)$	0.3981	No
	$m\,(0.9)$	0.0546	No**
	$m\,(0.1)$	0.6826	No
Year 2 vs. Year 3	$m\,(0.5)$	0.8746	No
	$m\,(0.9)$	0.0245	No*
	$m\,(0.1)$	0.0998	No**
Year 1 vs. Year 3	$m\,(0.5)$	0.2693	No
	$m\,(0.9)$	0.4442	No

Note: "No" denotes no significant difference at 1%; "No*" denotes difference significant at 5% but not at 1%; "No**" denotes difference significant at 10% but not at 1% nor 5%.

training in statistics.

To keep things in perspective, we report in Table 4.5 how well the classical prediction fairs against the evidence.

Since the absolute values of the t-statistics in Table 4.5 are large relative to the critical values (Table 4.1), it follows that the classical prediction is strongly rejected for students trained in statistics.

Table 4.4: t-test results.

Group	Matching probability	Mean	Standard deviation	Sample size	t-stat	Sig diff
	$m\,(0.1)$	0.1786	0.1296	23	0.2789	No
Econ	$m\,(0.5)$	0.4080	0.1706	23	-0.2449	No
	$m\,(0.9)$	0.7603	0.2315	20	1.2531	No
	$m\,(0.1)$	0.1871	0.1429	236	1.7275	No
Non-econ	$m\,(0.5)$	0.4126	0.1712	239	-0.3662	No
	$m\,(0.9)$	0.7227	0.2036	243	2.0881	No
	$m\,(0.1)$	0.1829	0.1362	116	0.9371	No
Male	$m\,(0.5)$	0.4239	0.1738	120	0.4557	No
	$m\,(0.9)$	0.7319	0.2046	124	1.9838	No
	$m\,(0.1)$	0.1892	0.1462	143	1.4846	No
Female	$m\,(0.5)$	0.4023	0.1682	142	-1.0181	No
	$m\,(0.9)$	0.7199	0.2071	139	1.3919	No
	$m\,(0.1)$	0.1773	0.1465	171	0.5579	No
Year 1	$m\,(0.5)$	0.4204	0.1729	173	0.2838	No
	$m\,(0.9)$	0.7195	0.2136	172	1.4763	No
	$m\,(0.1)$	0.1817	0.0939	27	0.5893	No
Year 2	$m\,(0.5)$	0.4070	0.1357	29	-0.3837	No
	$m\,(0.9)$	0.7820	0.1799	28	2.5457	No
	$m\,(0.1)$	0.2150	0.1576	60	2.1601	No
Year 3	$m\,(0.5)$	0.3909	0.1816	59	-1.0900	No
	$m\,(0.9)$	0.7134	0.1930	62	0.7323	No
	$m\,(0.1)$	0.2076	0.1555	126	2.638	Yes (1%)
Stat	$m\,(0.5)$	0.4043	0.1653	128	-0.847	No
	$m\,(0.9)$	0.7398	0.1710	130	2.957	Yes (1%)
	$m\,(0.1)$	0.1663	0.1243	133	-0.441	No
Non-stat	$m\,(0.5)$	0.4198	0.1762	134	0.206	No
	$m\,(0.9)$	0.7117	0.2344	133	0.800	No

Table 4.5: Comparison with classical probabilities.

Matching probability	Mean	Standard deviation	Sample size	Classical probability	t-stat	Sig diff
$m\,(0.1)$	0.2076	0.1555	126	0.1	7.767	Yes
$m\,(0.5)$	0.4043	0.1653	128	0.5	-6.550	Yes
$m\,(0.9)$	0.7398	0.1710	130	0.9	-10.682	Yes

4.9 Summary and conclusions

In this paper, we reported the results of our tests of the matching probabilities predicted by the quantum model of al-Nowaihi and Dhami (2017). These predicted matching probabilities agreed with those we observed (section 4.7). According to our Mann-Whitney U-tests, none of the demographic characteristics were significant. The only demographic characteristic we found to be significant according to our t-tests was a prior training in statistics (section 4.8). However, even for these students, the quantum prediction is far closer to the evidence than the classical prediction.

We showed (Proposition 4.1) that the Ellsberg paradox reemerges if we combine the heuristic of insufficient reason and behavioral assumption of this paper (section 4.3) with classical (non-quantum) probability theory. Hence, unlike earlier quantum models, this paper makes essential use of quantum probability.

Our derivation is parameter free, recall (4.4). Thus our model is more parsimonious than any of the alternatives. Our model is in accord with stylized facts 1 (*insensitivity*), 2 (*exchangeability*) and 3 (*no error*), see section 4.2. At the end of section 4.2, we suggested it may also be in accord with stylized facts 4 (*salience*) and 5 (*anonymity*); however, this could be a topic for future research.

Appendix A

to Chapter 2

A.1 Proofs

Proof of Proposition 2.1: 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 (2.14)-(2.20), it follows that the utility of a player who is a member of APR1 can be written as*
$U_1^{APR}(g_1, g_2, \theta_1) = \Phi_1^{APR}(g_1, \theta_1) + \Psi_1^{APR}(g_2)$, *where*
$\Phi_1^{APR}(g_1, \theta_1) = v_1(y - g_1) + r_1 g_1$
$+\nu_1 \left\{ \alpha_1 \left[\int_{x=0}^{g_1} (g_1 - x) f_1^2(x|\theta_1) dx \right] - \beta_1 \left[\int_{x=g_1}^{y} (x - g_1) f_1^2(x|\theta_1) dx \right] \right\}$
$+ (1 - \nu_1) \left\{ \alpha_1 \left[\int_{x=0}^{g_1} (g_1 - x) f_1^4(x) dx \right] - \beta_1 \left[\int_{x=g_1}^{y} (x - g_1) f_1^4(x) dx \right] \right\}$,
is a function of g_1, θ_1 but not of g_2; and
$\Psi_1^{APR}(g_2) = r_1 g_2$
$+\mu_1 \left\{ \gamma_1 \left[\int_{x=0}^{g_2} (g_2 - x) f_1^1(x) dx \right] - \delta_1 \left[\int_{x=g_2}^{y} (x - g_2) f_1^1(x) dx \right] \right\}$
$+ (1 - \mu_1) \left\{ \gamma_1 \left[\int_{x=0}^{g_2} (g_2 - x) f_1^3(x) dx \right] - \delta_1 \left[\int_{x=g_2}^{y} (x - g_2) f_1^3(x) dx \right] \right\}$,
is a function of g_2 but not of g_1.
Analogous expressions hold for members of APR2 and PUB.

Proof of Proposition 2.2: 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 maximise $U_1^{APR}(g_1, g_2, \theta_1)$ 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:*

$$\frac{\partial}{\partial g_1} U_1^{APR}(g_1, g_2, \theta_1) = r_1 + \beta_1 - v_1'(y - g_1) + (\alpha_1 - \beta_1) \left[\nu_1 F_1^2(g_1|\theta_1) + (1 - \nu_1) F_1^4(g_1) \right].$$

$$\frac{\partial^2}{\partial g_1^2} U_1^{APR}(g_1, g_2, \theta_1) = v_1''(y - g_1) + (\alpha_1 - \beta_1)\left[\nu_1 f_1^2(g_1|\theta_1) + (1-\nu_1) f_1^4(g_1)\right].$$

$$\frac{\partial^2}{\partial g_1 \partial \alpha_1} U_1^{APR}(g_1, g_2, \theta_1) = \nu_1 F_1^2(g_1|\theta_1) + (1-\nu_1) F_1^4(g_1).$$

$$\frac{\partial^2}{\partial g_1 \partial \beta_1} U_1^{APR}(g_1, g_2, \theta_1) = \nu_1\left[1 - F_1^2(g_1|\theta_1)\right] + (1-\nu_1)\left[1 - F_1^4(g_1)\right].$$

$$\frac{\partial^2}{\partial g_1 \partial \theta_1} U_1^{APR}(g_1, g_2, \theta_1) = \nu_1(\alpha_1 - \beta_1)\frac{\partial F_1^2(g_1|\theta_1)}{\partial \theta_1}.$$

Analogous expressions hold for APR2 (except that we do not condition on θ) and PUB (except that we have $F_1^4(g_1|\theta_1)$ instead of $F_1^4(g_1)$ and $f_1^4(g_1|\theta_1)$ instead of $f_1^4(g_1)$).

Proof of Proposition 2.3: 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. ■

Proof of Proposition 2.4: Similar to the proof of Proposition 2.5, below. ■

Proof of Proposition 2.5: 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 2.5(a) then follows from Lemma 3. Part (b) is similar. ■

Proof of Proposition 2.6: Suppose $\nu_1 = 1$, so that intentions are unimportant. From (2.19), and last line of (2.23) for $i = 1$, we see that $\phi_1^I(g_1) = \phi_1^I(g_1, \theta_1) = 0$. Hence, from (2.14), and (2.21) for $i = 1$, we see that the choice relevant terms in $U_1^{APR}(g_1, g_2, \theta_1)$ and $U_1^{PUB}(g_1, g_2, \theta_1)$ are the same. Since, by assumption, guilt-aversion is more important than surprise-seeking ($\alpha_1 \leq \beta_1$), it follows, from Proposition 2.3, that the optimal contribution (in each case) is unique. Hence, $\widehat{g}_1 = g_1^*$. ■

A.2 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.[102] 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.

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

[102]Note for the reader: These correspond to Stages-1 and Stage-2 in subsection 2.5.1.

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[103]

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.

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

[103]Note for the reader: This corresponds to Stages-1 in subsection 2.5.1. 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.

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[104]

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.

[104]Note for the reader: These were the instructions for the first part (corresponds to Stages-1 in subsection 2.5.1) 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[105]

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[106]

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.

[105] Note for the reader: This corresponds to Stage-2 in subsection 2.5.1. These are the instructions for the public treatment in our experiment. The private and public treatment were run in a counterbalanced order.

[106] 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).[107]

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

A.3 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.

A.4 More on psychological utility functions

Recall, from subsection 2.3.5, 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 2.3.5. 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 2.2 but for first order beliefs of player 1 in the PUB treatment.

Example A.1. *: 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, $i = 1, 2$.*

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

We consider the public treatment (PUB). Player 1 *has a first order belief about the contribution,* g_2*, 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:*

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

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

Geometrically, the density (A.1), (A.2) 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 (A.1) 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 (A.2) 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 (A.1) and (A.2) are, respectively,*

$$F_1^1(x|\theta_2) = \frac{x^2}{2\theta_2},\ x \in [0,\theta_2],\ \theta_2 \in (0,2]. \tag{A.3}$$

$$F_1^1(x|\theta_2) = \frac{2x - \frac{1}{2}x^2 - \theta_2}{2-\theta_2},\ x \in [\theta_2,2],\ \theta_2 \in [0,2). \tag{A.4}$$

A large number (in fact, an infinite number) of unconditional distributions are consistent with (A.1)-(A.4). For example, let player 1*'s prior distribution of* θ_2 *(before he sends the signal containing a realization of* θ_2*) be:*

$$\pi_1^1(\theta_2) = 1 - \frac{1}{2}\theta_2, \theta_2 \in [0,2], \tag{A.5}$$

According to (A.5), player 1 *believes that the most probable contribution of player* 2 *is zero. But many other prior distributions are consistent with (A.1)-(A.4), includ-*

ing:

$$\pi_1^1(\theta_2) = \frac{1}{2}\theta_2, \theta_2 \in [0,2], \tag{A.6}$$

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

$$f_1^1(x) = \int_{\theta=0}^{\theta=2} f_1^1(x|\theta)\,\pi_1^1(\theta)\,d\theta, \tag{A.7}$$

then (A.5), along with (A.1) and (A.2), imply the unconditional density:

$$f_1^1(0) = 0,\ f_1^1(x) = (\ln 2)\,x - x\ln x,\ x \in (0,2], \tag{A.8}$$

and, hence, the unconditional cumulative distribution:

$$F_1^1(0) = 0,\ F_1^1(x) = \frac{1}{4}x^2 + \frac{1}{2}(\ln 2)\,x^2 - \frac{1}{2}x^2\ln x,\ x \in (0,2]. \tag{A.9}$$

Of course, had we used (A.6) instead of (A.5), in conjunction with (A.1), (A.2) and (A.7), we would have got unconditional distributions different from (A.8) and (A.9).

A.4.1 Psychological utility for the APR treatment

Recall that the psychological utility function of a player 1 was given by (2.14) in subsection 2.3.5. It is now given by (A.10), below, and the *psychological utility function* of a player 2 in APR2 is now given by (A.11), below that.

$$\begin{aligned} U_1^{APR}(g_1, g_2, \theta_1) &= u_1(g_1, g_2) + \psi_1^S(g_2) + \phi_1^S(g_1, \theta_1) \\ &\quad + \psi_1^I(g_2) + \phi_1^I(g_1), \end{aligned} \tag{A.10}$$

$$\begin{aligned} U_2^{APR}(g_2, g_1) &= u_2(g_2, g_1) + \psi_2^S(g_1) + \phi_2^S(g_2) \\ &\quad + \psi_2^I(g_1) + \phi_2^I(g_2). \end{aligned} \tag{A.11}$$

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 (A.10), depends on θ_1 but the utility of player 2, in (A.11), does not depend on a signal.

Note that $\psi_1^S(g_2)$, $\psi_1^I(g_2)$ in (A.10) 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 (2.14) in subsection 2.3.5 without affecting any of the results. Similar remarks apply to the two functions $\psi_2^S(g_1)$, $\psi_2^I(g_1)$ in (A.11). 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 (2.14) and (2.14) that do affect choices. Let

$$\mu_i \in [0,1], \gamma_i \geq 0, \delta_i \geq 0, i = 1, 2, \tag{A.12}$$

these complement the parameters in (2.16).Consider the function $\psi_1^S(g_2)$ in (A.10). 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,

$$\psi_1^S(g_2) = \mu_1 \left\{ \gamma_1 \left[\int_{x=0}^{g_2} (g_2 - x) f_1^1(x)\, dx \right] - \delta_1 \left[\int_{x=g_2}^{y} (x - g_2) f_1^1(x)\, dx \right] \right\}. \tag{A.13}$$

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 (A.11) is defined by

$$\psi_2^S(g_1) = \mu_2 \left\{ \gamma_2 \left[\int_{x=0}^{g_1} (g_1 - x) f_2^1(x)\, dx \right] - \delta_2 \left[\int_{x=g_1}^{y} (x - g_1) f_2^1(x)\, dx \right] \right\}. \tag{A.14}$$

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.[108] 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 (2.14) and (2.17) of subsection 2.3.5 and retained in (A.10) above. Analogously for $\phi_2^S(g_2)$ in (2.15) and (2.18) of subsection 2.3.5 and retained in (A.11) 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.

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

Now, consider the function $\psi_1^I(g_2)$ in (A.10) above, and (A.15) below.

$$\begin{aligned}\psi_1^I(g_2) = (1-\mu_1)\Bigg\{&\gamma_1\left[\int_{x=0}^{g_2}(g_2-x)\,f_1^3(x)\,dx\right]\\ &-\delta_1\left[\int_{x=g_2}^{y}(x-g_2)\,f_1^3(x)\,dx\right]\Bigg\}.\end{aligned} \tag{A.15}$$

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 (A.15), 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).[109] 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 $\psi_1^I(g_2) < 0$.[110] Thus, we call $\psi_1^I(g_2)$ the *intentional surprise function for player* 1. Analogously, $\psi_2^I(g_1)$, in (A.11) above, and (A.16) below, we call *the intentional surprise function for player* 2.

$$\begin{aligned}\psi_2^I(g_1) = (1-\mu_2)\Bigg\{&\gamma_2\left[\int_{x=0}^{g_1}(g_1-x)\,f_2^3(x)\,dx\right]\\ &-\delta_2\left[\int_{x=g_1}^{y}(x-g_1)\,f_2^3(x)\,dx\right]\Bigg\}.\end{aligned} \tag{A.16}$$

We now give an argument to motivate $\phi_1^I(g_1)$ in (2.14) and (2.19) of subsection 2.3.5 and retained in (A.10), 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

[109] This makes sense because we do not require consistency of action and beliefs, see Section 2.3.4 above.

[110] 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.

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 (2.15) and (2.20) of subsection 2.3.5 and retained in (A.11) above.

A.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:

$$\begin{aligned} U_i^{PUB}(g_i, g_{-i}, \theta_i, \theta_{-i}) &= u_i(g_i, g_{-i}) + \psi_i^S(g_{-i}, \theta_{-i}) + \phi_i^S(g_i, \theta_i) \\ &\quad + \psi_i^I(g_{-i}, \theta_{-i}) + \phi_i^I(g_i, \theta_i), \end{aligned} \tag{A.17}$$

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

$$\begin{aligned} \psi_i^S(g_{-i}, \theta_{-i}) = \mu_i \Bigg\{ &\gamma_i \left[\int_{x=0}^{g_{-i}} (g_{-i} - x) f_i^1(x|\theta_{-i})\, dx \right] \\ &- \delta_i \left[\int_{x=g-i}^{y} (x - g_{-i}) f_i^1(x|\theta_{-i})\, dx \right] \Bigg\}, \end{aligned} \tag{A.18}$$

$$\begin{aligned} \psi_i^I(g_{-i}, \theta_{-i}) = (1 - \mu_i) \Bigg\{ &\gamma_i \left[\int_{x=0}^{g_{-i}} (g_{-i} - x) f_i^3(x|\theta_{-i})\, dx \right] \\ &- \delta_i \left[\int_{x=g_{-i}}^{y} (x - g_{-i}) f_i^3(x|\theta_{-i})\, dx \right] \Bigg\}, \end{aligned} \tag{A.19}$$

and the parameters are as in (A.12) above.

The interpretation of (A.17), (A.18) and (A.19) is the same as (A.10) to (A.16) except for the introduction of the conditioning on θ_i, θ_{-i}.

Appendix B

to Chapter 3

B.1 Proofs

Proof of Proposition 3.1: Using (3.23), in the domain, $e \geq b_W^2$, we have $\frac{dU}{de} < 0$ so that the optimal effort in this domain takes its lowest value, $b_W^2 = e^*$. Now consider the domain $e < b_W^2$. In the first row of (3.23), the first two terms are negative, while the third is non-negative. There are the following cases to consider:

(i) Let $Y_G < e - Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right)$ for all levels of effort in the interval $[0, b_W^2)$, then $\frac{dU}{de} < 0$ at all points. A sufficient condition for this to hold is $Y_G < -Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right)$. In this case, $\frac{dU}{de} < 0$ in both domains, $e < b_W^2$ and $e \geq b_W^2$, so the optimal effort choice is the lowest possible, $e^* = 0$.

(ii) Suppose that $Y_G > e - Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right)$ for all levels of effort in the interval $[0, b_W^2)$, then $\frac{dU}{de} > 0$ at all points. Then the optimal effort is the highest possible effort in the interval $[0, b_W^2)$. A sufficient condition is $Y_G > b_W^2 - Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right)$.

(iii) Suppose that $-Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right) \leq Y_G \leq b_W^2 - Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right)$. In this case, U is not monotonic in the interval $[0, b_W^2)$. Given the concavity of U we have an interior solution in the interval $[0, b_W^2)$. Setting the first order condition equal to zero, we have $e^* = Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right) + Y_G < b_W^2$. This is also the global optimum because corresponding to all effort levels greater than e^* we have $\frac{dU}{de} < 0$. Using (3.5), we also directly get the result that $Y_R(\theta)$ is highest in the Kind treatment (so effort is the highest) and lowest in the Unkind treatment (so effort is the lowest); the result for the case of the Neutral treatment is intermediate. ■

Proof of Corollary 1: Suppose that the second order beliefs increase from b_W^2 to $b_W^2 + \varepsilon$, where $\epsilon > 0$. If $Y_G < -Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right)$, as in Proposition 3.1(i), then after the increase in beliefs, we still have $e^* = 0$. Similarly in the intermediate case in Proposition 3.1(iii), there is no change in optimal effort. The only change occurs if $Y_G > b_W^2 - Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right)$. In this case, there are two possibilities.

(1) $Y_G > b_W^2 + \varepsilon - Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right)$, so the optimal solution is at $e^* = b_W^2 + \varepsilon$.

(2) $Y_G < b_W^2 + \varepsilon - Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right)$. In this case, the optimal solution $e^* \in (b_W^2, b_W^2 + \varepsilon)$.

In each of the two cases, there is an increase in optimal effort. Overall, across all cases, effort should be increasing in second order beliefs. ■

Proof of Proposition 3.2: We separately examine the behavior of U in the two intervals $[0, b_W^2)$ and $[b_W^2, \overline{e}]$.

I. The domain $e \geq b_W^2$.

Note from (3.23) that over the interval $[b_W^2, \overline{e}]$, $\frac{dU}{de} \gtreqless 0$ iff $Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right) \gtreqless e$. Thus, $Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right) > \overline{e}$, or $Y_R > \frac{\overline{e}}{(p-w)\left(w - \frac{1}{2}\overline{w}\right)}$, is a sufficient condition for $\frac{dU}{de} > 0$ over the interval $[b_W^2, \overline{e}]$. In this case, the optimal solution in $[b_W^2, \overline{e}]$ is $e^* = \overline{e}$.

On the other hand, $Y_R(p-w)\left(w-\frac{1}{2}\overline{w}\right) < b_W^2$, or $Y_R < \frac{b_W^2}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$, is a sufficient condition for $\frac{dU}{de} < 0$ over the interval $[b_W^2, \overline{e}]$. In this case, the optimal solution in $[b_W^2, \overline{e}]$ is $e^* = b_W^2$. It follows that when $\frac{b_W^2}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)} < Y_R < \frac{\overline{e}}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$ we have an interior solution in $[b_W^2, \overline{e}]$, which can be found by setting the first order condition to zero, and this gives $e^* = Y_R(p-w)\left(w-\frac{1}{2}\overline{w}\right)$.

II. The domain $e < b_W^2$.

In the domain $[0, b_W^2)$, (3.23) gives $\frac{dU}{de} \gtreqless 0$ iff $Y_R(p-w)\left(w-\frac{1}{2}\overline{w}\right) \gtreqless e - Y_G$. It follows that $Y_R > \frac{b_W^2 - Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$, is a sufficient condition for $\frac{dU}{de} > 0$ over $[0, b_W^2)$ and the optimal solution is the highest effort level in the interval $[0, b_W^2)$. Since $Y_R > 0$ and $w - \frac{1}{2}\overline{w} \geq 0$, the case $Y_R < \frac{-Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$ which ensures $\frac{dU}{de} < 0$ over the entire interval $[0, b_W^2)$ is ruled out. Thus, we cannot have a corner solution at $e = 0$. It follows that when $Y_R < \frac{b_W^2 - Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$ the only possibility is that we have an interior solution that can be found by setting the first order condition equal to zero:

$$e^* = Y_R(p-w)\left(w-\frac{1}{2}\overline{w}\right) + Y_G < b_W^2 \tag{B.1}$$

(i) Let $Y_R > \frac{\overline{e}}{(w-b_W^1)}$. This also implies that $Y_R > \frac{b_W^2 - Y_G}{(w-b_W^1)}$. Hence, $\frac{dU}{de} > 0$ in both intervals $[0, b_W^2)$ and $[b_W^2, \overline{e}]$. It follows that the optimal effort choice is the highest possible, $e^* = \overline{e}$.

(ii) Let $\frac{b_W^2}{(w-b_W^1)} < Y_R < \frac{\overline{e}}{(w-b_W^1)}$. From the results above, we have the following. In the domain $[b_W^2, \overline{e}]$, the condition $\frac{b_W^2}{(w-b_W^1)} < Y_R < \frac{\overline{e}}{(w-b_W^1)}$ implies that the optimal solution in this domain is $e^* = Y_R(w - b_W^1)$. Since $\frac{b_W^2}{(w-b_W^1)} \leq Y_R$ we also have that $Y_R > \frac{b_W^2 - Y_G}{(w-b_W^1)}$. In the domain $[0, b_W^2)$, the condition $\frac{b_W^2 - Y_G}{(w-b_W^1)} < Y_R$ implies that the optimal solution is the highest effort level in the interval $[0, b_W^2)$ and U is increasing throughout this interval. Thus, the globally optimal solution is $e^* = Y_R(w - b_W^1)$. Straightforward differentiation and the use of (3.5) gives the comparative static results.

(iii) Let $\frac{b_W^2 - Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)} < Y_R < \frac{b_W^2}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$. From the results above, we have the following. In the domain $[b_W^2, \overline{e}]$, the condition $Y_R < \frac{b_W^2}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$ implies that the optimal solution in this domain is $e^* = b_W^2$. In the domain $[0, b_W^2)$, the condition $\frac{b_W^2 - Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)} < Y_R$ implies that the optimal solution is the highest effort level in the interval $[0, b_W^2)$. These conditions imply that in the domain $[0, b_W^2)$, U is increasing throughout and in the domain $[b_W^2, \overline{e}]$ it is falling throughout. Hence, the optimal is $e^* = b_W^2$.

(iv) Let $Y_R \leq \frac{b_W^2 - Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$, which also implies that $Y_R < \frac{b_W^2}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$. In the domain $[b_W^2, \overline{e}]$ we have that $\frac{dU}{de} < 0$ so the optimal solution in this domain is $e^* =$

b_W^2. In the domain $[0, b_W^2)$, the only possibility is that we have an interior solution, e^*, given by (B.1). Since U is strictly concave, to the right of e^* is strictly decreasing for all points in the domain $(e, \overline{e}]$. Hence, the global maximum in this case is e^*, given by (B.1). As in the proof of Proposition 3.1(iii), we have the highest effort in the Kind treatment, the lowest in the Unkind treatment and intermediate in the Neutral treatment. ■

Proof of Corollary 2: The proof follows by considering all the cases in Proposition 3.2. Suppose that the initial level of second order beliefs are given by any feasible arbitrary value b_W^2 and the final level is give by $b_W^2 + \varepsilon$, where $\varepsilon > 0$ is defined separately in each case below. We show that the optimal effort level cannot decrease at the new level of beliefs.

(a) If $Y_R > \frac{\overline{e}}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$, then $e^* = \overline{e}$ (Proposition 3.2(i)). All quantities here are independent of b_W^2, so optimal effort cannot be reduced when b_W^2 increases.

(b) Let $\frac{b_W^2}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)} < Y_R < \frac{\overline{e}}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$ so that the optimal solution is $e^* = Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right) > b_W^2$ (Proposition 3.2(ii)). Then there exists some $\varepsilon > 0$ such that $\frac{b_W^2+\varepsilon}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)} < Y_R < \frac{\overline{e}}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$ and effort continues to be $e^* = Y_R(p - w)\left(w - \frac{1}{2}\overline{w}\right) > b_W^2 + \varepsilon$.

(c) If $\frac{b_W^2-Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)} < Y_R < \frac{b_W^2}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$, then the optimal solution is $e^* = b_W^2$ (Proposition 3.2(iii)). There exists some ε such that $\frac{b_W^2+\varepsilon-Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)} < Y_R < \frac{b_W^2+\varepsilon}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$. In this case, the optimal solution is $e^* = b_W^2 + \varepsilon$, so there is an increase in optimal effort.

(d) If $Y_R \leq \frac{b_W^2-Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$, then the optimal solution is $e^* = Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right) + Y_G < b_W^2$ (Proposition 3.2(iv)). For any $\varepsilon > 0$ such that $Y_R \leq \frac{b_W^2+\varepsilon-Y_G}{(p-w)\left(w-\frac{1}{2}\overline{w}\right)}$, in which case the optimal solution is $e^* = Y_R(p-w)\left(w - \frac{1}{2}\overline{w}\right) + Y_G < b_W^2 + \varepsilon$, which is unchanged.

Overall, it follows across all cases that effort is non-decreasing. Since the initial choice of beliefs was arbitrary, we expect a positive correlation in the data between effort and second order beliefs. ■

Proof of Proposition 3.4: (i) Let $w - b_W^1 \leq 0$. From (3.25), $\frac{dV}{de} < 0$ so optimal effort is $e^{G1} = 0$.

(ii) Let $w - b_W^1 > 0$, then $\frac{dV}{de}$ evaluated at $e = 0$ is strictly positive, hence, $e^{G1} > 0$. Solving out for e^{G1} from the first order condition, we get $e^{G1} = \gamma(\theta)(w - b_W^1)(p - w)$ if $\gamma(\theta)(w - b_W^1)(p - w) < \overline{e}$, otherwise optimal e^{G1} equals $\overline{e}$. If $e^{G1} = \gamma(\theta)(w - b_W^1)(p - w)$, then successively differentiating both sides with respect to w and b_W^1 we get

$$\frac{de^{G1}}{dw} = (p - w) - (w - b_W^1) \gtreqless 0.$$

$$\frac{de}{db_W^1} = -(p - w) < 0.$$

The result follows by using (3.20) in $e^G = \gamma(\theta)(w - b_W^1)(p - w)$. ■

B.2 A note on the optimal choice of wage

The focus of our paper is on the explanation of effort choice by workers. Here we offer a few comments on a separate issue that is not the focus of the paper–the choice of optimal wage by the firm.

If the firm behaves as a Stackelberg leader, as in classical game theory, then it maximizes π_F, given in (3.1), by a suitable choice of w_1, w_2, subject to the effort reaction function of the workers. However, the effort reaction function depends on which model (gift exchange or guilt–aversion/reciprocity) describes the worker's preferences. In the case of the gift exchange model, the effort reaction function is given in Proposition 3.4. In the case of guilt–aversion/reciprocity, it is given by Propositions 3.1 (negative reciprocity) and 3.2 (positive reciprocity). In each case, the effort reaction function is highly non-linear and the endpoints of the intervals themselves depend on the firm's choice of a wage, w.

The effort reaction functions depend on the first order beliefs of the workers, b_W^1. These are not observed by the firms who will need to form second order beliefs about b_W^1. But these second order beliefs are not observable to the experimenter, which creates enormous difficulties for empirical testing.

Consider, for instance, the simplest case in which the worker uses the gift exchange model and this is known to the firm. Proposition 3.4 gives the effort reaction function. Suppose that the firm believes that the distribution of b_W^1 is given by H. With probability $H(w)$ we have that b_W^1 is less than w, and the worker is in the domain of positive gift exchange, so effort is given by (3.26). With the complementary probability, effort is zero (Proposition 3.4(i)), so profits are also zero. The firm then chooses w to maximize its expected profit for worker, $i = 1, 2$, given by

$$E\pi_{Fi} = H(w)(p - w)^2(w - b_W^1)$$

Even in the simplest case of a uniform H, $E\pi_{Fi}$ is a cubic function of w. Furthermore, we do not observe H and even if we were to find the optimal w by simulating around a hypothetical H, the determinants of w are the parameters of this distribution. The analysis with reciprocal/guilt-averse firms and imperfect information about which model the worker uses is far more complex. In this case, the firm must form estimates of the second order beliefs of the worker by using its third order beliefs that are non-observable.

For these reasons we defer the firm's problem to future research.

B.3 Experimental instructions

Instructions for all participants

You are now participating in an economics experiment. If you read all the instructions carefully, you will be able to earn some money – depending on your decisions and others' decisions. Therefore it is important to actually read the instructions very carefully. During the experiment it is 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, our currency will not be 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 the exchange rate of **2 tokens = 1 Yuan**. Additionally, you will receive 5 Yuan, as a show-up fee for participating in this experiment.

There are two types of participants in this experiment: firms and workers. Each firm will be grouped with two workers– worker 1 and worker 2. Your role (firm or worker) is randomly chosen. You don't know who the other persons in your group are. The experiment consists of only one period with two stages.

Background information[111]:

(Kind treatment) In the past, the firm in your group not only approved workers' paid leave but also extended it by two days.

(Unkind treatment) In the past, the firm in your group found an excuse to put off workers' paid leave.

Stage 1:

Firms pay wages, which in total should not exceed their endowment (200 tokens); and each worker's wage is restricted to the multiples of 10 tokens upto a 100 tokens, i.e. $\{0, 10, ..., 100\}$. Firms are required to guess the possible effort level for each of the two workers, i.e., the effort levels they expect the workers to choose(from the set $\{0.1, 0.2, ..., 1.0\}$. Normally, workers in previous experiments undertook the effort level of 0.4.

[111]Note to the reader: There is no background information in the Neutral treatment.

Workers are required to state their possible effort level from the set $\{0.1, 0.2, ..., 1.0\}$. This possible effort is just a willingness of the effort level the workers would like to choose subsequently. Only the worker him/herself knows his/her own possible effort, and other players in the experiment cannot obtain this information. Workers are also required to guess the firm's possible wage paid to them, i.e., the wage they expect the firm to choose from the set $\{0, 10, ..., 100\}$.

Additionally, each firm and worker may have a chance to win an additional 5 Yuan as a prize. At the end of the experiment, we will randomly choose one firm and one worker from those whose guess was correct. If nobody guessed accurately, then we will randomly choose from the ones whose guess is the closest to the partner's choice and give him/her a 2 Yuan prize. If the firm's expectations for both workers are correct, then the probability of winning the prize is higher.

Stage 2:

Workers observe their own wage and choose their actual effort level from the set $\{0.1, 0.2, ..., 1.0\}$. Additionally, the choices of actual effort will be conveyed to the grouped firm in the end. The following revenues are calculated using this actual effort.

Calculation of revenues:

1. The firm's revenue is calculated as follows:

(100 − worker 1's wage)×worker 1's actual effort + (100 − worker 2's wage)×worker 2's actual effort

2. Each worker receives the wage paid by the firm, net of the cost of the actual effort level chosen. The cost of effort is shown below:

Effort	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Cost	0	1	2	4	6	8	10	12	15	18

The lowest effort (0.1) costs the worker nothing. The cost of effort increases with the effort level, as does the revenue produced for the firm. The increment in costs also increases with the effort level. Worker's revenue is calculated as follows:

worker's wage − cost of the actual effort

Control questions

These questions are hypothetical. Please answer questions 1 and 2 in terms of *tokens*.

1. A firm paid 20 tokens in wage to each worker. Worker 1 exerted the effort 0.2, while, worker 2 exerted the effort 0.6. What are the revenues for. . . ?

Firm: Worker 1: Worker 2:

2. A firm paid 50 tokens in wage to worker 1 and 60 tokens in wage to worker 2. Worker 1 exerted the effort 0.3, while worker 2 exerted the effort 0.7. What are the revenues for…?

Firm: Worker 1: Worker 2:

3. If you were a worker and you guessed a correct wage, then you may win __ Yuan additional prize;

If no worker guessed correctly, then the one who guessed most closely will win __ Yuan additional prize.

4. If you were a firm and you guessed the correct effort, then you may win __ Yuan additional prize;

If no firm guessed correctly, then the one who guessed most closely will win __ Yuan additional prize.

Post-experiment survey

We would like to ask you a few questions.

1. Age: __ years

Gender: (female/male)

Field of study: __

Degree of study: __

Year of study: __

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

2. (*Firm*) How did you form the expectations of workers' efforts?

A. You used the effort level in previous experiments (0.4) as a reference.

B. The extension (or refusal) of the worker's paid leave in the background information influenced your expectations. (Please specify)

C. You used information other than in A and B to predict worker's choice. (Please specify)

2. (*Workers*) How did you form the expectation of wage?

A. You used the effort level in previous experiments (0.4) as a reference.

B. The extension (or refusal) of worker's paid leave in the background information influenced your expectations. (Please specify)

C. You used information other than in A and B to predict the firm's choice. (Please specify)

Appendix C

to Chapter 4

C.1 Experimental Instruction (translation from Chinese instruction)

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. You will receive 5 Yuan for participation. This is irrespective of your decisions in the experiment. 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. The experiment will be carried out only once.

This experiment is paper based. there are three tasks: Task 1, Task 2 and Task 3. In each task, there are two boxes- Box K and Box U, and each box contains 100 colored balls. The composition of the balls is known for Box K but unknown for Box U. After you complete a task, the experimenter will collect the materials for that task and you will receive the materials for the next task.

Task 1:

There are 50 *purple* balls and 50 *yellow* balls in Box K. For each of the eleven rows in Table 1, tick *exactly one* of the following boxes: "Receive x Yuan for sure", "Indifferent" or "Play Box K".

Box U contains 100 balls (purple or yellow) but in *unknown* proportions. Thus Box U can contain any number of purple balls from 0 to 100 and any number of yellow balls from 0 to 100 provided the sum of balls (purple plus yellow) is 100. The composition of Box U will be randomly decided at the end of the experiment. For each of the eleven rows in Table 2, tick *exactly one* of the following boxes: "Receive x Yuan for sure", "Indifferent" or "Play Box U".

In each table, if you believe that you are indifferent between the choice in the left column and the right column, you may tick the box under the middle column "Indifferent".

At the end of the experiment, one of the eleven rows of Table 1 or one of the eleven rows of Table 2 will be selected at random and played for real money. In Table 1, you will receive x Yuan for sure if you have ticked the box under "Receive x Yuan for sure" or, if you have ticked the box under "Play Box K", you will win 10 *Yuan* if a *purple* ball is drawn from Box K (otherwise you win nothing). In Table 2, you will receive x Yuan for sure if you have ticked the box under "Receive x Yuan for sure" or, if you have ticked the box under "Play Box U" you will win 10 Yuan if a *purple* ball is drawn from Box U (otherwise you win nothing). In each table, if you have ticked "Indifferent" in the randomly selected row, then one of the left or right cells in this selected row will be randomly chosen to play for real.

Table 1		
Receive x Yuan for sure	Indifferent	Play Box K
$x = 10$ ☐	☐	☐
$x = 9$ ☐	☐	☐
$x = 8$ ☐	☐	☐
$x = 7$ ☐	☐	☐
$x = 6$ ☐	☐	☐
$x = 5$ ☐	☐	☐
$x = 4$ ☐	☐	☐
$x = 3$ ☐	☐	☐
$x = 2$ ☐	☐	☐
$x = 1$ ☐	☐	☐
$x = 0$ ☐	☐	☐

Table 2		
Receive x Yuan for sure	Indifferent	Play Box U
$x = 10$ ☐	☐	☐
$x = 9$ ☐	☐	☐
$x = 8$ ☐	☐	☐
$x = 7$ ☐	☐	☐
$x = 6$ ☐	☐	☐
$x = 5$ ☐	☐	☐
$x = 4$ ☐	☐	☐
$x = 3$ ☐	☐	☐
$x = 2$ ☐	☐	☐
$x = 1$ ☐	☐	☐
$x = 0$ ☐	☐	☐

After you complete Task 1, the experimenter will collect the materials for Task 1 and you will receive the materials for Task 2.

Task 2:

There are 100 balls of 10 different colors (including *purple*) in Box K. There are exactly 10 balls of each color. For each of the eleven rows in Table 1, tick *exactly* *one* of the following boxes: "Receive x Yuan for sure", "Indifferent" or "Play Box K".

Box U contains 100 balls of the same colors as in Box K but in *unknown* proportions. Thus, Box U could contain any number of purple balls from 0 to 100. And similarly for each of the other 9 colors (provided the sum of balls of all colors is 100). The composition of Box U will be randomly decided at the end of the experiment. For each of the eleven rows in Table 2, tick *exactly one* of the following boxes: "Receive x Yuan for sure", "Indifferent" or "Play Box U".

In each table, if you believe that you are indifferent between the choice in the left column and the right column, you may tick the box under the middle column "Indifferent".

At the end of the experiment, one of the eleven rows of Table 1 or one of the eleven rows of Table 2 will be selected at random and played for real money. In Table 1, you will receive x Yuan for sure if you have ticked the box under "Receive x Yuan for sure". However, if you have ticked "Play Box K", then you shall win 10 *Yuan* if a *purple* ball is drawn from Box K (otherwise you win nothing). In Table 2, you will receive x Yuan for sure if you have ticked the box "Receive x Yuan for sure". However, if you have ticked the box "Play Box U" then you win 10 *Yuan* if a *purple* ball is drawn from Box U (otherwise you win nothing). In each table, suppose that you ticked "Indifferent" in the randomly selected row, then one of the left or right cells in this selected row will be randomly chosen to play for real.

After you complete Task 2, the experimenter will collect the materials for Task 2 and you will receive the materials for Task 3.

Task 3:

As in task 2, there are 100 balls in Box K of 10 different colors (including *purple*). There are exactly 10 balls of each color. For each of the eleven rows in Table 1, tick *exactly one* of the following boxes: The box "Receive x Yuan for sure", "Indifferent" or "Play Box K".

As with task 2, Box U contains 100 balls of the same colors as in Box K but in *unknown* proportions. For each of the eleven rows in Table 2, tick *exactly one* of the following boxes: The box "Receive x Yuan for sure", "Indifferent" or "Play Box U".

In each table, if you believe that you are indifferent between the choice in the left column and the right column, you may tick the box under the middle column "Indifferent".

At the end of the experiment, one of the eleven rows of Table 1 or one of the eleven rows of Table 2 will be selected at random and played for real money. In

Table 1, you will receive x Yuan for sure if you tick the box under "Receive x Yuan for sure". However, now if you have ticked "Play Box K", then you shall win 10 *Yuan* if a *non-purple* ball is drawn from Box K (otherwise you win nothing). In Table 2, you will receive x Yuan for sure if you tick the box "Receive x Yuan for sure". However, if you have ticked the box "Play Box U" then you win 10 *Yuan* if a *non-purple* ball is drawn from Box U (otherwise you win nothing). In each table, suppose that you tick "Indifferent" in the randomly selected row, then one of the left or right cells in this selected row will be randomly chosen to play for real.

After you have completed Task 3, the experimenter will collect the materials for Task 3 and the experiment will terminate.

C.2 Post-experimental Questionnaire

1. Age: ____ years old
2. Gender: (female/male)
3. Field of study: ______
4. Degree of study: ______
5. Year of study: ______
6. Have you participated in similar experiments in the past? (Yes/No)
7. Did you have statistics course(s) before? (Yes/No)
8. Does your preference of some particular color(s) affect your decisions?

A. No. B. Yes. Please specify how your preference of some particular color(s) affected your decisions below.

Bibliography

[1] Abdellaoui M. A., Baillon A., Placido L. and Wakker P. P. (2011). The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation. *The American Economic Review*. 101(2): 695–723.

[2] Akerlof, G. A. (1982). Labor Contracts as Partial Gift Exchange. *The Quarterly Journal of Economics*. 97(4): 543–69.

[3] Asano M., Basieva I., Khrennikov A., Ohya M., & Tanaka Y. (2016). A Quantum-like Model of Selection Behavior. *Journal of Mathematical Psychology*. http://dx.doi.org/10.1016/j.jmp.2016.07.006.

[4] Aerts D. and Aerts S. (1994). Applications of Quantum Statistics in Psychological Studies of Decision Processes. *Foundations of Science*. 1: 85–97.

[5] al-Nowaihi A. and Dhami S. (2015). Evidential Equilibria: Heuristics and Biases in Static Games of Complete Information. *Games*. 6: 637–676.

[6] al-Nowaihi A. and Dhami, S. (2017). The Ellsberg Paradox: A Challenge to Quantum Decision Theory? *Journal of Mathematical Psychology*. http://dx.doi.org/10.1016/j.jmp.2016.08.003.

[7] al-Nowaihi A., Ian Bradley I. and Dhami S. (2008). The Utility Function Under Prospect Theory. *Economics Letters* 99: 337–339.

[8] Basieva I. and Khrennikov A. (2015). On the Possibility to Combine the Order Effect with Sequential Reproducibility for Quantum Measurements. *Foundations of Physics*. 45(10): 1379–1393.

[9] Battigalli P. and Dufwenberg M. (2007). Guilt in Games. *The American Economic Review*. 97(2): 170–176.

[10] Battigalli P. and Dufwenberg M. (2009). Dynamic Psychological Games. *Journal of Economic Theory*. 144(1): 1–35.

[11] Bell D E. (1985). Disappointment in Decision Making under Uncertainty. *Operations research*. 33(1): 1–27.

[12] Benartzi S. and Thaler R. H. (1995). Myopic Loss Aversion and the Equity Premium Puzzle. *The Quarterly Journal of Economics*. 110(1): 73–79.

[13] Berg J., Dickhaut J., McCabe K. (1995). Trust, Reciprocity, and Social History. *Games and economic behavior*. 10(1): 122–142.

[14] Bicchieri C. (2006). The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge: Cambridge University Press.

[15] Busemeyer J. R. and Bruza P. D. (2012). *Quantum Models of Cognition and Decision*. Cambridge University Press.

[16] Camerer C. (2003). Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press: Princeton.

[17] Camerer C., Thaler R. H. (1995). Anomalies: Ultimatums, Dictators and Manners. *The Journal of Economic Perspectives*. 9(2): 209–219.

[18] Chalmers A. F. (1999). *What Is This Thing Called Science?* Third edition. Open University Press.

[19] Charness G. and Dufwenberg M. (2006). Promises and Partnership. *Econometrica*. 74(6): 1579–1601.

[20] Charness G., Frechette G. R., Kagel J. H. (2004). How Robust Is Laboratory Gift Exchange? *Experimental Economics*. 7(2): 189–205.

[21] Charness G. and Kuhn P. (2007). Does Pay Inequality Affect Worker Effort? Experimental Evidence. *Journal of Labor Economics*. 25(4): 693–723.

[22] Chow C. C. and Sarin R. K. (2001). Comparative Ignorance and the Ellsberg Paradox. *Journal of Risk and Uncertainty*. 22: 129–139.

[23] Chow C. C. and Sarin R. K. (2002). Known, Unknown and Unknowable Uncertainties. *Theory and Decision*. 52: 127–138.

[24] Croson R. and Uri G. (2009). Gender Differences in Preferences. *Journal of Economic Literature*. 47(2): 1–27.

[25] Curley S. P., Yates J. F. and Abrams R. A. (1986). Psychological Sources of Ambiguity Avoidance. *Organizational Behavior and Human Decision Processes*. 38: 230–256.

[26] Dhami, S. (2016). *The Foundations of Behavioral Economic Analysis*. Oxford University Press: Oxford.

[27] Dhami, S., Wei, M. and al-Nowaihi, A. (2016) Public Goods Games and Psychological Utility: Theory and Evidence. University of Leicester Working Paper Series No. 16/17.

[28] Dimmock S. G., Kouwenberg R. and Wakker P. P. (2015). Ambiguity Attitudes in a Large Representative Sample. *Management Science*. Article in Advance. Published Online: November 2, 2015. 10.1287/mnsc.2015.2198.

[29] Dufwenberg M., Gächter, S. and Hennig-Schmidt, H. (2011). The Framing of Games and the Psychology of Play. *Games and Economic Behavior*. 73(2): 459–478.

[30] Dufwenberg M. and Gneezy U. (2000). Measuring Beliefs in an Experimental Lost Wallet Game. *Games and Economic Behavior*. 30: 163–182.

[31] Dufwenberg M. and Kirchsteiger G. (2000). Reciprocity and Wage Undercutting. *European Economic Review*. 44(4): 1069–1078.

[32] Dufwenberg M. and Kirchsteiger G. (2004). A Theory of Sequential Reciprocity. *Games and Economic Behavior*. 47(2): 268–298.

[33] Eckel C. C. and Grossman P. J. (2008). Differences in the Economic Decisions of Men and Women: Experimental Evidence. In Handbook of Experimental Economics Results, Volume 1, ed. Charles Plott and Vernon Smith, (pp. 509-519). New York: Elsevier.

[34] Ellingsen T., M. Johannesson, S. Tjøtta, and G. Torsvik (2010). Testing Guilt Aversion. *Games and Economic Behavior*. 68: 95–107.

[35] Ellsberg D. (1961). Risk, Ambiguity and the Savage Axioms. *The Quarterly Journal of Economics*. 75: 643–669.

[36] Ellsberg D. (2001). *Risk, Ambiguity and Decision*. Garland Publishers, New York. Original Ph.D. dissertation: Ellsberg D. 1962. Risk, Ambiguity and Decision. Harvard, Cambridge, MA.

[37] Elster J. (1989). Social Norms and Economic Theory. *Journal of Economic Perspectives*. 3(4): 99–117.

[38] Elster J. (1998). Emotions in Economic Theory. *Journal of Economic Literature*. 36: 47–74.

[39] Englmaier F. and Leider S. (2012). Contractual and Organizational Structure with Reciprocal Agents. *American Economic Journal: Microeconomics.* 4(2): 146–183.

[40] Falk A., Fehr E., Fischbacher U. (2008). Testing Theories of FairnessIntentions Matter. *Games and Economic Behavior.* 62(1): 287–303.

[41] Fehr E., Gächter S. and Kirchsteiger G. (1997). Reciprocity as a Contract Enforcement Device: Experimental Evidence. *Econometrica.* 65(4): 833–860.

[42] Fehr E., Kirchsteiger G. and Riedl A. (1993). Does Fairness Prevent Market Clearing? An Experimental Investigation. *The Quarterly Journal of Economics.* 108(2): 437–459.

[43] Fehr E., Kirchsteiger, G. and Riedl A. (1998). Gift Exchange and Reciprocity in Competitive Experimental Markets. *European Economic Review.* 42(1): 1–34.

[44] Fehr E., Naef M., and Schmidt K. M. (2006). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment. *The American Economic Review.* 96(5): 1912–1917.

[45] Fehr E. and Schmidt K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics.* 114(3): 817–68.

[46] Fehr E. and Schmidt K. M. (2006). The Economics of Fairness, Reciprocity and Altruism: Experimental Evidence and New Theories. in Serge-Christophe Kolm and Jean Mercier Ythier (eds.) Handbook of the Economics of Giving, Altruism and Reciprocity, Volume 1., Elsevier.

[47] Fischbacher U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental economics.* 10(2): 171–178.

[48] Fox C. R. and Tversky A. (1995). Ambiguity Aversion and Comparative Ignorance. *The Quarterly Journal of Economics.* 110(3): 585–603.

[49] Gächter S., Nosenzo D., Sefton M. (2013). Peer Effects in Pro-social Behavior: Social Norms or Social Preferences? *Journal of the European Economic Association.* 11(3): 548–573.

[50] Geanakoplos J., Pearce D. and Stacchetti E. (1989). Psychological Games and Sequential Rationality. *Games and Economic Behavior.* 1(1): 60–79.

[51] Gnedenko B. V. (1968). *The Theory of Probability.* Forth edition. Chelsea Publishing Company, New York, N. Y.

[52] Gneezy U. and Aldo R. (2004). Gender and Competition at a Young Age. *The American Economic Review*. 94(2): 377–381.

[53] Guerra G. and Zizzo D. J. (2004). Trust Responsiveness and Beliefs. *Journal of Economic Behavior and Organization*. 55: 25–30.

[54] Harrison G. W. (1986). An Experimental Test for Risk Aversion. *Economics Letters*. 21: 7–11.

[55] Kahneman D. and Tversky A. (2000). *Choices, Values and Frames*. Cambridge University Press.

[56] Keynes J. M. (1921). *A Treatise on Probability*. London: Macmillan Co.

[57] Khalmetski K. (2016). Testing Guilt Aversion with an Exogenous Shift in Beliefs. *Games and Economic Behavior*. 97: 110–119.

[58] Khalmetski K., Ockenfels A., and Werner P. (2015). Surprising Gifts. *Journal of Economic Theory*. 159: 163–208.

[59] Khrennikov A., Basieva I., Dzhafarov E. N. and Busemeyer J. R. (2014). Quantum Models for Psychological Measurement: An Unsolved Problem. PLoS ONE. 9(10): e110909.

[60] Kolmogorov A. N. (1950). *Foundations of the Theory of Probability*, New York: Chelsea Publishing Co. Original German, 1933.

[61] Kothiyal A., Spinu V. and Wakker P. P. (2014). An Experimental Test of Prospect Theory for Predicting Choice under Ambiguity. *Journal of Risk and Uncertainty*. 48(1): 1–17.

[62] La Mura P. (2009). Projective Expected Utility. *Journal of Mathematical Psychology*. 53(5): 408–414.

[63] Levine D. K. (1998). Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics*. 1(3): 593–622.

[64] Loomes G., Sugden R. (1986). Disappointment and Dynamic Consistency in Choice under Uncertainty. *The Review of Economic Studies*. 53(2): 271–282.

[65] Malmendier U. and Schmidt K. (2017). You Owe Me. *The American Economic Review*. 107(2): 493–526.

[66] Prelec D. (1998) The Probability Weighting Function. *Econometrica*. 60: 497–524.

[67] Pulford B. D. and Colman A. M. (2008). Ambiguity Aversion in Ellsberg Urns with Few Balls. *Experimental Psychology*. 55(1): 31–37.

[68] Rabin M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review*. 83(5): 1281–1302.

[69] Reuben E., Sapienza P. and Zingales L. (2009). Is Mistrust Self-fulfilling? *Economics Letters*. 104(2): 89–91.

[70] Ross L., Greene D. and House P. (1977). The 'False Consensus Effect': An Egocentric Bias in Social Perception and Attribution Processes. *Journal of Experimental Social Psychology*. 13(3): 279–301.

[71] Shapiro C. and Stiglitz J. E. (1984). Equilibrium Unemployment as a Worker Discipline Device. *The American Economic Review*. 74(3): 433–444.

[72] Thaler R. H. (1999). Mental Accounting Matters. *Journal of Behavioral Decision Making*. 12: 183–206.

[73] Tolman R. C. (1938). *The Principles of Statistical Mechanics*. Oxford University Press, Oxford.

[74] Trautmann S. T., Vieider F.M. and Wakker P. P. (2008). Causes of Ambiguity Aversion: Known versus Unknown Preferences. *Journal of Risk and Uncertainty*. 36(3): 225–243.

[75] Tversky A. and Kahneman D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*. 5: 297–323.

[76] Vanberg C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica*. 76(6): 1476–1480.

[77] von Neumann J. (1955). *Mathematical Foundations of Quantum Theory*. Princeton Press. Original German, 1932.

[78] Wooldridge J. M. (2015). *Introductory Econometrics: A Modern Approach*. Nelson Education.