CeDEx

Centre for Decision Research and Experimental Economics

Discussion Paper Series

ISSN 1749-3293

Discussion Paper

CeDEx Discussion Paper No. 2006–20

The Framing of Games and the Psychology of Strategic Choice

Martin Dufwenberg, Simon Gächter and Heike Hennig-Schmidt

October 2006

The University of Nottingham

CeDEx

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of Public Economics, Individual Choice under Risk and Uncertainty, Strategic Interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/economics/cedex/ for more information about the Centre or contact

Karina Whitehead
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0) 115 95 15620
Fax: +44 (0) 115 95 14159
karina.whitehead@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/economics/cedex/papers/index.html

The Framing of Games and the Psychology of Strategic Choice

Martin Dufwenberg, Simon Gächter & Heike Hennig-Schmidt

October 16, 2006

Abstract:
Psychological game theory can provide a rational choice explanation of framing effects; frames influence beliefs, and beliefs influence motivations. We explain this point theoretically, and explore its empirical relevance experimentally. In a 2×2-factorial framing design of one-shot public good experiments we show that frames affect subject's first- and second-order beliefs, and contributions. From a psychological game-theoretic framework we derive two mutually compatible hypotheses about guilt aversion and reciprocity under which contributions are related to second- and first-order beliefs, respectively. Our results are consistent with either.

Keywords: Framing; psychological games; guilt aversion; reciprocity; public good games; voluntary cooperation

JEL codes: C91, C72, D64, Z13.

Acknowledgments: This paper is part of the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms. Support from the EU-TMR Research Network ENDEAR (FMRX-CT98-0238) and the University of St. Gallen Grundlagenforschungsfonds is gratefully acknowledged. Part of the research was done when Martin Dufwenberg visited Bonn University on a F. W. Bessel Research Award from the Alexander-von-Humboldt Foundation. We thank Christian Thöni for very helpful research assistance. We are also grateful for comments from Paolo Battigalli, Gary Charness, Eva Poen, Arthur Schram, Martin Sefton, Jean-Robert Tyran, Manuela Vieth, and participants of the Sozialwissenschaftlicher Ausschuss of the Verein für Socialpolitik, the Gesellschaft für Experimentelle Wirtschaftsforschung, the Midlands Game Theory Workshop, and workshop participants at the Max Planck Institute of Economics in Jena.

Contact: Martin Dufwenberg, University of Arizona, martind@eller.arizona.edu; Simon Gächter, University of Nottingham, IZA and CESifo, simon.gaechter@nottingham.ac.uk; Heike Hennig-Schmidt, University of Bonn, hschmidt@uni-bonn.de

1. INTRODUCTION

Experimental evidence from psychology and economics has shown that the framing of decisions may matter to preferences and choice (cf. Pruitt 1967, Selten and Berg 1970). This may reflect a failure by decision makers to exhibit "elementary requirements of consistency and coherence", as found by Tversky and Kahneman (1981) in a classic paper. Our main objective is to theoretically articulate, and experimentally illustrate, a further reason why framing may matter. We shall make no reference to irrationality, and the sort of framing effect we highlight comes from an interaction of frames and players' motivation. Framing may influence strategic behavior in games by influencing motivation that depends on beliefs, about choices and beliefs, in subtle ways.

Our message partly echoes the insight that focal points may influence coordination, as first noted by Schelling (1960) and investigated experimentally by Mehta, Starmer and Sugden (1994). The idea here is that a description of a strategic situation may possess cues that serve to coordinate choice behavior. This entails that descriptions influence *beliefs about others' choices*, which in turn may have bearing on a person's rational choice. However, we push beyond this observation as follows: We argue that if players are emotional or care for the intentions and desires of others, then framing may influence behavior independently of how beliefs about others' choices change. Frames may influence *beliefs about others' beliefs*, which in itself may influence a person's choice even if his or her belief about others' actions is given. The reason is that if players are emotional or care for the intentions and desires of others, then motivations may depend on beliefs directly.

The upshot is that framing may play a very special role in *psychological games*, as defined by Geanakoplos, Pearce and Stacchetti (1989). These structures differ from standard games in that the domain of a payoff function includes beliefs, not just strategy profiles. A body of recent work (cited in more detail below) in experimental economics and behavioral theory argues that psychological games are needed to capture some important 'social' preferences, like reciprocity or guilt aversion (a desire not to let others down).

In psychological games, motivation depends on beliefs (about choices and about beliefs) directly, so if beliefs are changed motivation may flip too. The key contribution of this paper is to tie this observation in with framing effects: frames may influence beliefs, which, as we just said, spells action in psychological games. Effectively, what we propose is an explanation why framing may matter when decision-makers interact strategically.

The paper has two main parts. First, in Section 2, we provide a theoretical elucidation regarding the potential relevance of our new approach to framing effects. Second, in Section 3, we report the results of an experiment designed to explore the empirical significance of the idea. We choose a simple public good game as our workhorse, and derive and test predictions based on two psychological-game based models. Our findings here may be of independent interest to the experimental literature on public goods (and social dilemma) games. Section 4 concludes.

2. FRAMING EFFECTS IN PSYCHOLOGICAL GAMES

The key idea of this paper is that frames may shape players' beliefs in games, which may in turn influence strategic choices. Part of this message is reflected in the literature on focal points, which goes back to Schelling (1960). Schelling noted that in many games certain choices are 'focal', which may facilitate coordination. A classic example involves two persons meeting in New York City: going to Grand Central Station may be a focal choice.

Schelling's NYC example involves focal points created by properties possessed by a particular strategy, but one can easily imagine how focal points are similarly created by the framing of a game. Consider the two following games, which differ only by name:

The let's get 7 game:

	a	***b***
a	(9, 9)	(0, 8)
b	(8, 0)	(7, 7)

The let's get 9 game:

	a	***b***
a	(9, 9)	(0, 8)
b	(8, 0)	(7, 7)

Are these games the same? They have player sets, strategy sets, and payoff functions in common. However, the games' names differ, and different names may trigger different beliefs in the minds of the players. The players may, for example, coordinate on different equilibria in the two cases: Imagine that each player chooses strategy ***b*** in the first game, while each player chooses ***a*** in the second game. This would illustrate how frames could, in principle, serve as equilibrium selection devices.[1]

In the preceding example, frames shape beliefs and beliefs influence behavior, ultimately because of what beliefs tell a player about a co-player's choices. So far so good, but nothing in this is really original. The new thing in this paper is instead to point out that the link from frames to beliefs to actions does not necessarily rely on perceptions of others' behavior. To make this point as clear as possible, consider the following example. The example will exhibit how a frame may influence a player's beliefs which influence the player's behavior, and yet it is from the outset inconceivable that any other player's behavior could change.

The example concerns the dictator game, which involves two players:[2] The first player, the *dictator*, chooses how to divide a sum of money, say $1000, between the two players. The second player, the *recipient*, has no real choice—she has to simply *accept* the dictator's decision.

[1] The let's get 7 and let's get 9 games are so-called stag-hunt games, amply discussed for the intriguing coordination problem embodied. This matters *e.g.* to theories of equilibrium selection (*e.g.* Harsany & Selten 1988; Carlsson & van Damme 1993), examinations of the impact of communication (*e.g.* Aumann 1990; Charness 2000, Clark, Kay and Sefton 2001), and the impact of learning (*e.g.* Crawford 1995). We thus add framing to this list of topics.

[2] Forsythe, Horowitz, Savin and Sefton (1994) were among the first to study the dictator game, which has subsequently been used in many experimental studies.

Now assume that the dictator does not like to let others down; he suffers from guilt (an emotional response) if he gives others less than he believes they expect. We will say that the dictator is *guilt averse*.[3]

Now imagine that one ran an experiment on this game, with the twist of calling it by different names in different treatments. Say that the game was referred to as either

- *The let's split a grand game*, or
- *The German tipping game*.

Imagine that most subjects make the equal split in the first case, that most subjects give away just small change in the second case, and that all of this happens because the dictator subjects hold vastly different beliefs about what recipients expect to get in the two cases. Under the first frame, dictators choose the equal split because that is what they expect recipients to expect them to do, and the dictators would feel exceedingly guilty unless they lived up to these expectations. Under the second frame, they give away peanuts because this is all they expect the recipient to expect. This illustrates how a frame could, in principle, influence a dictator's beliefs, which influence his motivation, which influences his behavior, despite there being no strategic uncertainty whatsoever about what other players do.

We would like to make two more comments about this example. The first is that the example makes reference to a form of motivation which is non-standard, in the sense that what is going on cannot be modeled using traditional game theory. To see this, note that in traditional theory of (normal form) games any player i has a utility of the form

$$u_i: A \to \mathbb{R}, \qquad (1)$$

[3] The terminology is adapted from Charness and Dufwenberg (2006). Dufwenberg and Gneezy (2000) report experimental dictator game evidence in line with guilt aversion, although they do not consider framing effects.

where A is the set of strategy profiles of the game. Applied to the dictator game, $A = A_{\text{dictator}} \times A_{\text{recipient}}$ = [\$0, \$1000]×{*accept*}, where A_{dictator} and $A_{\text{recipient}}$ are the players' respective strategy sets (the elements of the former specifying how much the dictator gives away). Such a formulation, whether used to model selfishness or some kind of other-regarding motivation (like altruism, or inequity aversion) predicts a uniquely defined set of best responses for *i*.[4] By contrast, in the example, the guilt averse dictator's set of best responses depends on his beliefs about the recipient's beliefs. Hence (1) cannot describe those preferences; hence traditional game theory is not a rich enough toolbox to handle this case.

To model belief-dependent preferences, such as the dictator's guilt aversion in the example, one must move to utilities of the following form:

$$u_i: A \times M_i \to \mathbb{R}, \qquad (2)$$

where M_i is *i*'s beliefs (about choices and beliefs), somehow described. This means that we need to move from standard games to so-called psychological games, as introduced by Geanakoplos *et al*.[5]

Our second comment relates to (2) and the extent to which that utility specification is actually more general than is borne out by the example involving the guilt averse dictator. That example showed how a frame may influence a player's beliefs, which influence his motivation, which influences his behavior, despite there being no strategic uncertainty whatsoever about what other players do. The formulation (2), however, is by no means limited to that case; (2) allows that, as frames change, so do beliefs of any order, including beliefs about others' strategies. What

---

[4] Examples: If the dictator is selfish, then his set of best responses is {\$0}; if his objective is to minimize the difference between his payoff and the recipients, then his set of best responses is {\$500}; if his objective is to maximize the maximum payoff to one of the players, then his set of best responses is {\$0, \$1000}.

[5] Cf. also Battigalli & Dufwenberg (2005) who generalize Geanakoplos *et al*'s framework in several directions (including allowing updated beliefs to influence utility and having incomplete information; such extension are important in many applications, but do not concern us here however).

comes out of this may be framing effects that have a hybrid quality to them: frames may influence a player's beliefs, which influence his motivation directly *as well as* his perception of others' choices, and all of this influences his behavior. In other words, beliefs about anything in the domain of the utility in (2) may move behavior, and the domain includes *both* others' choices and others' beliefs. All those, potentially complex, links from frames to beliefs to actions are what we have in mind when we talk about our new approach to framing effects.

3. FRAMING & FREE-RIDING: AN EXPERIMENT

In this section we report the results of an experiment, designed to examine the empirical relevance of the ideas introduced in Section 2. The subsections that follow will in turn:

A. introduce, and motivate the choice of, our vehicle of research: a public good game,
B. discuss framing issues,
C. incorporate guilt aversion and reciprocity using psychological game theory,
D. derive testable predictions for the thus derived psychological public good games,
E. present the experimental design and procedures,
F. report the results.

3.A *The public good game*

As a vehicle of investigation we wish to select a game for which framing effect have already been documented, so that we can relate to and help further understand previous work. A linear public good game is our choice. Numerous experiments have shown the existence of framing effects (cf. section 3.B), and since public good games represent many economically important situations that require the agents' voluntary cooperation it is important to understand how frames affect voluntary cooperation.

Linear public good games have the advantage of being simple, which makes a psychological game-theoretic analysis tractable (cf. sections 3.C-D). The simplicity is due to the fact that selfish players have a dominant strategy to free ride, *i.e.*, subjects' optimal behavior is independent of others' behavior. Yet, numerous experiments have shown that many people do not play accordingly (Ledyard 1995). In particular, although previous work has not made connection to psychological games, there is evidence that subjects' choices may depend on their beliefs.[6] Thus, since our argument is that frames may influence beliefs and beliefs may affect motivations and thereby behavior, public good experiments are well-suited for our purposes.

We consider a public good game with the following structure: Each of three players simultaneously chooses how to allocate twenty monetary units between a 'private' and a 'public' account. The sum of what the players contribute to the public account is multiplied by 1.5, to determine its total value. A player's earnings is the sum of whatever he or she puts in the private account, plus one third of the total value of the public account.

The situation can be represented as a normal form game $G=(A_i, \pi_i)_{i\in N}$ such that $N=\{1,2,3\}$ is the player set, $A_i = \{0,1,\ldots,20\}$ is the strategy set of player i, and $\pi_i\colon \times_{j\in N} A_j \rightarrow \mathbb{R}$ is i's monetary payoff function defined by

$$\begin{aligned}\pi_i(a_1, a_2, a_3) &= 20 - a_i + (1/3)\cdot(3/2)\cdot(a_1 + a_2 + a_3) = \\ &= 20 - a_i + \tfrac{1}{2}\cdot(a_1 + a_2 + a_3). \qquad (3)\end{aligned}$$

All experimental treatments are set up to implement that structure. The treatments differ only in the frames used (cf. sections 3.B and 3.E).

[6] For instance, Croson (2002) and Fischbacher and Gächter (2006) have shown that a subject's contribution is often highly positively correlated with the subject's beliefs about others' contributions. This result is anticipatory of one of our hypotheses (H_2) below.

3.B *Framing*

In selecting the frames to be examined, we take inspiration from some previous work on framing in the context of social dilemma-type games. A common distinction concerns whether a frame changes a reference point or whether it just consists of different wordings. We refer to these as *valence framing* and *label framing*. The former is a description that puts the same essential information in either a positive or a negative light (Levin, Schneider and Gaeth 1998). Several studies have looked at valence framings in public good provision. In the standard public good experiment subjects are endowed with some money, which they can keep for themselves or contribute to the public good. We will call this situation a *give treatment*. Thus, any contribution to the public good is a positive externality for all other players by definition of a public good. Another framing is to endow the group with the resources and to allow the group members to withdraw resources; we will call this the *take treatment*.[7]

Label framing is involved if subjects are confronted with alternative, but objectively equivalent problem wordings (see, e.g., Elliott, Hayward and Canon 1998 who call this a "pure framing effect"). Ross and Ward (1996) and Liberman, Samuels and Ross (2004) report one of the best-known labeling effects. In their experiments a simple prisoners' dilemma game was either called the "Community Game", or the "Wall Street Game". Otherwise, the game and the instructions were identical. Cooperation rates were significantly lower under the "Wall Street" frame than under the "Community Game" frame.[8]

[7] The typical result from numerous experiments is that in the give treatment contributions to the public good are higher than in the take treatment (cf. Andreoni 1995; Sonnemans, Schram and Offerman 1998; Willinger and Ziegelmeyer 1999; Cookson 2000; Park 2000; Ostrom 2006; Brandts and Schwieren 2006).

[8] Rege and Telle (2004) similarly played a one-shot public goods experiment and found higher contributions under a "community" frame than under a neutral frame. Further studies on label frame effects comprise Burnham, McCabe, and Smith (2000) and Abbink and Hennig-Schmidt (2006).

Previous studies have either looked at value framing or at label framing, but we study both simultaneously. As we explain further in section 3.E, we develop a 2×2 design that varies both the label (community versus neutral treatments) and the valence (give versus take treatments). Thus, our factorial design allows us to assess the relative importance of these different versions of framing effects.

We also move beyond the existing literature on framing by eliciting first- and second-order beliefs. This will help us testing hypotheses based on psychological game theory which can embrace framing effects. We discuss this latter topic next.

3.C *Guilt aversion and reciprocity*

Most of economic theory depicts decision makers as 'selfish', in the sense that they care only about their own monetary payoffs. In the context of the public good game we consider, this would correspond to assuming that (3) (or possibly a similar formulation modified to control for risk-aversion) can describe the players' preferences. By contrast, a rich body of experimental evidence suggests that decision makers often have more complex objectives, and in particular that they somehow care about what others get or do or hope to achieve. Some theoretical models have been proposed, with the objective to model such social preferences, and some of the social preference models build on psychological game theory.[9] We focus on two of these – guilt aversion and reciprocity – and use them to derive testable implications that we subsequently address in the experiment.

[9] For discussions of the experimental evidence as well as of many social preference models, see Fehr and Gächter (2000), Camerer (2003, Ch. 2), and Sobel (2005). Battigalli & Dufwenberg (2005) survey models that use psychological game theory.

Guilt aversion is dislike of the guilt felt if one chooses so as to give others less than one expects them to expect. Charness and Dufwenberg (2006) introduce the term.[10] We move directly to incorporating guilt aversion to the game at hand: Let b_{ij} denote i's 'first-order belief' about j's choice (i,j=1,2,3; $i \neq j$); b_{ij} is the mean of a probability distribution i has over A_j. Let c_{iji} denote i's 'second-order belief' about b_{ji}; c_{iji} is the mean of a probability measure i has over the possible values of b_{ji}. One way to model guilt aversion is to assume that i suffers from guilt to the extent that he puts less in the public account than the average of what he believes his two co-players believe he puts in the public account. Formally, his utility function u_i* can be defined by

$$u_i^*(a_1, a_2, a_3, c_{iji}, c_{iki}) = 20 - a_i + ½\cdot(a_1 + a_2 + a_3) - \gamma_i \cdot \max\{0, (c_{iji} + c_{iki})/2 - a_i\} \qquad (4)$$

where i, j, k = 1, 2, 3; $i \neq j \neq k \neq i$, and where $\gamma_i \geq 0$ is a parameter measuring i's degree of guilt aversion. If $\gamma_i = 0$, (4) has the same RHS as (3) and $a_i = 0$ is a dominant strategy. If $0 < \gamma_i < ½$, the RHS of (4) changes, but $a_i = 0$ is still a dominant strategy. However, if $\gamma_i > ½$ very different possibilities come alive, as i's best response will be $a_i = (c_{iji} + c_{iki})/2$. In this case i's best response is belief dependent.

With reference to (1) and (2) in Section 2, note that (4) has the form u_i: $A \times M_i \to \mathbb{R}$ rather than u_i: $A \to \mathbb{R}$, since (4) includes beliefs in its domain. $G^*=(A_i, u_i^*)_{i \in N}$ is a psychological game.

Reciprocity is a desire to get even, to respond to perceived wrongdoings with revenge and to reward perceived kindness. Rabin (1993) developed a theory of reciprocity, which made the meaning of words like "kindness" precise. Rabin argues that kindness depends on what a player believes about others' choices, as this can capture a player's 'intentions'. Moreover, reciprocal

[10] Similar motivations, under different names, are considered by Huang and Wu (1994), Dufwenberg and Gneezy (2000), Bacharach, Guerra and Zizzo (2004), and Dufwenberg (2002) who also draws connections to work on guilt in social psychology including Baumeister, Stillwell and Heatherton (1994, 1995).

motivation depends in general on beliefs about kindness, and hence on beliefs about beliefs since kindness depends on beliefs. Psychological game theory is called for.

Rabin's objective is to call attention to two central qualitative aspects of reciprocity, and he restricts attention to two-player normal form games. Falk and Fischbacher (2006) and Dufwenberg and Kirchsteiger (2004) provide extensions that allow for more players (and which also consider extensive games). The following draws on the latter model.

Applied to our game, the utility of player i is given by $u_i{**}$ defined by

$$u_i{**}(a_1, a_2, a_3, b_{ij}, b_{ik}) = 20 - a_i + ½\cdot(a_1 + a_2 + a_3) + \rho_i\cdot(\kappa_{ij}\cdot\lambda_{iji} + \kappa_{ik}\cdot\lambda_{iki}) \quad (5)$$

where again $i, j, k = 1, 2, 3$; $i \neq j \neq k \neq i$, and where the last term is in special need of further explanation. All but the last terms in (5) capture how the agent cares for own income (cf. (3) and (4)). The last term captures how he is motivated by reciprocity: $\rho_i \geq 0$ is a constant measuring i's sensitivity to reciprocity; κ_{ij}, κ_{ik}, λ_{iji}, and λ_{iki} depend on i's choice or beliefs: κ_{ij} represents i's kindness to j – it is positive (negative) if i is kind (unkind); λ_{iji} represents i's belief about how kind j is to i – it is positive (negative) if i believes that j is kind (unkind). κ_{ik} and λ_{iki} have analogous interpretations. Equation (5) captures reciprocity by making it in i's interest to match the signs of κ_{ij} and λ_{iji}, and of κ_{ik} and λ_{iki}, *ceteris paribus*.

We need to calculate κ_{ij}, λ_{iji}, κ_{ik}, and λ_{iki}. This turns out to be (more) straightforward (than analogous calculations in many other games). Although in general games kindness depends on beliefs this is not the case in the public good game because there is a one-to-one link between a player's choice and his kindness. This is because, independently of the co-players' choices, there is a one-to-one link between a player's choice and his impact on the other players' monetary payoffs. Player i's kindness to j [or k] is the difference between what i actually gives to

j [or k] and the average of the maximum (=20) and minimum (=0) that i could give to j [or k]. We get $\kappa_{ij} = \kappa_{ik} = a_i$–10. To get λ_{iji}, note first that this is i's belief about $\kappa_{ji} = a_j$–10, so just replace a_j by b_{ij} in the RHS of that expression; instead of $\kappa_{ij} = a_i$–10 we get $\lambda_{iji} = b_{ij}$–10. Similarly, we get $\lambda_{iki} = b_{ik}$–10. All in all, we can re-write the RHS of (5) to get (5'):

$$20 - a_i + ½\cdot(a_1 + a_2 + a_3) + \rho_i\cdot[(a_i - 10)\cdot(b_{ij} - 10) + (a_i - 10)(b_{ik} - 10)] =$$
$$= 20 - a_i + ½\cdot(a_1 + a_2 + a_3) + \rho_i\cdot[(a_i - 10)\cdot(b_{ij} + b_{ik} - 20)]. \quad (5')$$

If $b_{ij}+b_{ik}-20 \leq 0$ (5') is maximized by a_i=0 regardless of ρ_i. The interpretation is that i does not consider j and k to be, on average, kind, so there is no reason for i to sacrifice payoff to help j and k. If $b_{ij}+b_{ik}-20 > 0$, then (5') is maximized by $a_i = 20$ if ρ_i is large enough and by $a_i = 0$ if ρ_i is small enough. (Besides these cases there are some additional combinations of b_{ij}, b_{ik}, and ρ_i that make i indifferent between all his strategies.) The formulation joins the above one on guilt aversion in that what is a best response depends on i's beliefs, although different beliefs matter this time.

With reference to (1) and (2) in Section 2, note that (5) (based on (5')) has the form u_i: $A\times M_i \rightarrow \mathbb{R}$ rather than u_i: $A \rightarrow \mathbb{R}$, since the utility function defined in (5) (based on (5')) includes beliefs in its domain. Hence $G^{**}=(A_i, u_i^{**})_{i\in N}$ is a psychological game.

3.D *Hypotheses*

Our experiment is set up to test the two theories presented in Section 3.C, and to check if/how framing matters alongside. Since the utility functions (4) and (5) (based on (5')) include beliefs in their domains, the theories can be directly tested if one observes beliefs. Our design allows us to elicit some beliefs that are relevant to this task.[11] We describe the procedure for belief elicitation in detail in section 3.E.

We formulate our hypotheses with reference to the choices and beliefs of individual players, rather than in terms of some equilibrium that would give predictions for all players jointly. If we focused on equilibria, we would run the risk of incorrectly rejecting a valid insight about motivation only because people did not coordinate well. Our approach is consistent with the theory development in Section 3.C, where we merely discussed properties of an individual player's best responses rather than equilibrium.

Hypothesis H_1 concerns guilt aversion. Recall that if $\gamma_i < ½$ then i's best response is $a_i = 0$ even if $(c_{iji}+ c_{iki})/2 > 0$; if $\gamma_i > ½$ then i's best response is to match his or her second-order beliefs: $a_i = (c_{iji}+ c_{iki})/2$.[12] Our design allows us to measure and observe the second-order beliefs $(c_{iji}+ c_{iki})/2$. However, γ_i is unobservable so we have to make a few assumptions to derive a testable prediction. If one took the theory dead serious, and if one had full confidence that our measure of $c_{iji}+ c_{iki}$ is without noise or error, then it would be natural to test whether $a_i \in \{0, (c_{iji}+ c_{iki})/2\}$, with $a_i = (c_{iji}+ c_{iki})/2$ whenever $a_i > 0$. In reality, we neither expect the theory to be exactly correct nor the beliefs to be measured completely without error (cf. footnote 15 below). Therefore, we are lead to consider the following somewhat weaker statement. It separates between subjects with $a_i = 0$ & $(c_{iji}+c_{iki})/2 > 0$, for whom we 'know' that $\gamma_i < ½$, and the others:

[11] Dufwenberg and Gneezy (2000), Bacharach, Guerra and Zizzo (2001), and Charness and Dufwenberg (2006) have previously measured beliefs in experiments, with the purpose of connecting to psychological game theory. It is furthermore interesting to note that Ross and Ward (1996, p 108), who conducted that influential framing study we discussed in section 3.B, made some remarks which may be taken to indicate their interest in our approach. They call for further research on how a label influences the way subjects feel they ought to play and how a label changes their expectations about how the other player would choose to play. The authors conjecture that a frame may even alter subjects' beliefs about how the other player would expect them to play.

[12] If $\gamma=½$ then anything is a best response for i but we ignore this possibility.

H_1: $a_i = 0$ & $(c_{iji}+c_{iki})/2 > 0$ ***or*** there is a positive correlation between a_i and $(c_{iji}+c_{iki})/2$.

H_1 predicts that for subjects that contribute non-zero amounts contributions and second-order beliefs are positively correlated. To determine if we can support H_1, we perform a one-sided test of the null hypothesis of zero correlation between a_i and $(c_{iji}+c_{iki})/2$ considering only those i for which it is *not* the case that $a_i = 0$ & $(c_{iji}+c_{iki})/2 > 0$.

Hypothesis H_2 concerns reciprocity. Our design allows us to measure and observe the first-order beliefs b_{ij} and b_{ik}. However, as was the case with γ_i for the case of guilt aversion, ρ_i is unobservable so again we have to make a few assumptions to derive a testable prediction. Recall from section 3.C that with reciprocal motivation the predicted choices are either 0 [implied when $(b_{ij} + b_{ik})/2 \leq 10$] or 20 [which would imply $(b_{ij} + b_{ik})/2 > 10$].[13] Again allowing some leeway in the extent to which one takes the theory and accuracy of measured beliefs dead serious, it may seem natural to test whether there is a positive correlation between contributions a_i and the average of first-order beliefs $(b_{ij} + b_{ik})/2$. However, one can plausibly add a proviso concerning subjects who exhibit $a_i = 0$ and $(b_{ij} + b_{ik})/2 = 20$ we 'know' that ρ_i is so low that they would never reciprocate kindness with kindness. The following hypothesis treats these subjects separately (cf. the "$a_i = 0$ & $(c_{iji}+c_{iki})/2 > 0$" part of H_1):

H_2: $a_i = 0$ & $(b_{ij}+b_{ik})/2 = 20$ ***or*** there is a positive correlation between a_i and $(b_{ij}+b_{ik})/2$.

[13] We ignore the possibility that (for certain combinations of ρ_i, b_{ij}, and b_{ik}) i may be indifferent between all his choices. This time the assumption may not be quite as innocuous as the analogous assumption in the case of guilt aversion (cf. footnote 12) since the indifferences would not solely depend on the exogenous parameter ρ_i (in analogy to the $\gamma=½$ case) but also on the first-order beliefs b_{ij}, b_{ik} which would be endogenously determined if we applied some equilibrium concept. However, as explained in the text, we do not apply any equilibrium concept.

To determine if we can support H_2, we perform a one-sided test of the null of zero correlation between a_i and $(b_{ij}+b_{ik})/2$ considering only those i for which it is *not* the case that $a_i = 0$ & $(b_{ij}+b_{ik})/2 = 20$.

Note that hypotheses H_1 and H_2 differ with respect to the beliefs involved. Guilt-aversion operates via second-order beliefs whereas reciprocity works on first-order beliefs. Note also that our experiment is not set up to test guilt aversion against reciprocity; in our game the two theories do not necessarily imply mutually inconsistent testable predictions.

H_1 and H_2 represent directional research hypotheses derived from specific theories, and so will be submitted to one-sided tests. In addition to H_1 and H_2 we will examine framing effects, *i.e.* whether choices and first- and second-order beliefs differ by treatment. Here we have no theory to guide us, and hence we perform two-sided tests.

Note how all these tests are connected. This is a paper about how frames affect choices. We propose to understand this as a two-part linkage: (i) frames move beliefs, and (ii) beliefs shape motivation and choice. Guilt aversion and reciprocity theory entail specific statements about (ii); as regards (i) we have no theory and merely record what we see.

3.E *Experimental design*

The standard linear public good game (Ledyard 1995) as introduced in section 3.A is our workhorse. The subjects are randomly assigned to groups of three people and each subject is endowed with 20 'Taler' (the experimental currency).

We employ a 2×2 factorial design, which consists of two label and two valence frames. The label frame involves a minimal change in wording, naming the game in two different ways. In the NEUTRAL labeling, whenever the instructions or the decision screens refer to the experiment we speak of "the experiment". In the COMMUNITY labeling, whenever we refer to the experiment we name it "the community experiment".[14]

The valence frame entails describing the game as a "give-some" or a "take-some" game. The GIVE frame corresponds to the standard public good setting given in equation (3). The instructions explain carefully that (i) the Talers the subject keeps for herself generate an "income from Taler kept"; (ii) the Talers the subject contributes to a project of her group create an "income from the project"; (iii) the subject's total income is the sum of both kinds of income.

In the TAKE frame, subjects can take Talers from a "project", the public good. The parameters were chosen to make the monetary payoff function in the TAKE frame equivalent to the GIVE situation. Therefore, the project consists of 60 Talers. Each subject i can take $t_i \in \{0,1,...,20\}$ Talers from the project and the payoff function under the TAKE frame is given by

$$\pi_i(t_1, t_2, t_3) \qquad = t_i + ½\cdot(60 - (t_1 + t_2 + t_3)) \qquad (6)$$

Note that (6) describes the same monetary payoff function as (3), since $t_i = 20 - a_i$.

Table 1 summarizes our 2×2 design.

Table 1: *Our* 2×2-*design – experimental treatments*

Treatment name	Valence frame	Label Frame	Independent observations
GIVE-NEUTRAL	GIVE	NEUTRAL	66
GIVE-COMMUNITY	GIVE	COMMUNITY	51
TAKE-NEUTRAL	TAKE	NEUTRAL	72
TAKE-COMMUNITY	TAKE	COMMUNITY	66

[14] The name of the game was changed at four place in the instructions, once on the decision screen for contributing to (taking from) the project, twice on each of the decision screens for first and second order belief elicitation.

We ran the experiments in the Bonn Laboratory of Experimental Economics. All experiments were computerized, using the software z-Tree (Fischbacher 1999). In total, 255 people participated, almost all undergraduate students from Bonn University majoring in law, economics and other disciplines. We conducted 15 sessions (four in each of GIVE-NEUTRAL, TAKE-NEUTRAL and GIVE-COMMUNITY and three in TAKE-COMMUNITY) with 18 or 15 participants, respectively.

The above public good problem was explained to the subjects in the instructions (see Appendix). We took great care to ensure that subjects understood the game and the incentives. After subjects had read the instructions, for which they had plenty of time, they had to answer ten control questions that tested their understanding of the decision situation in the different treatment conditions. We did not proceed until all subjects had answered all questions correctly.

After subjects had answered the control questions, they had to make their contribution or take decision. We then asked them to guess, on the one hand, the sum of their co-players' contributions, and, on the other hand, the sum of their co-players' guesses. For each of these guesses subjects were paid €20 each if their guesses were exactly correct, and nothing otherwise. These guesses form the basis of our measurement of $(b_{ij}+ b_{ik})/2$ and $(c_{iji}+ s_{iki})/2$.[15] When subjects made their contribution or take decisions they did not know about the subsequent estimation tasks. We decided on this timing of events because we did not want subjects' choices of contributions to be influenced by what choice they thought might facilitate correct subsequent guesswork. Subjects played the game only *once* without being informed about their income before the end of the experiment. Thus, all decisions are strictly independent.

[15] The incentives provided do not exactly provide incentives for means-revelation, as would seem relevant to the theory in section 3.C. We chose our belief-elicitation protocol because it is simple and easy to explain. Our idea is to get a rough-but-meaningful ballpark estimate of the participants' first- and second-order beliefs.

We recruited subjects by campus advertisements that promised a monetary reward for participation in a decision-making task. In each session, subjects were randomly allocated to the cubicles, where they took their decisions in complete anonymity from the other participants. All participants were informed fully on all features of the experimental design and the procedures. Sessions lasted for about 1 hour. On average subjects earned €15.20 (roughly $15 at the time of the experiment).

3.F *Results*

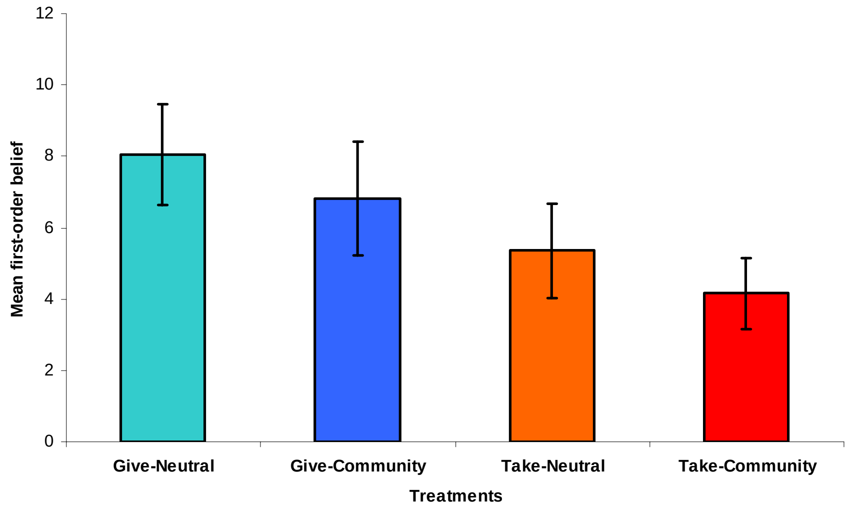
Recall the discussion, at the end of section 3.D, of how we propose to test a couple of hypotheses derived from theory (*i.e*, H_1 and H_2). In addition, our 2×2 design allows for a systematic but open-minded search for framing effects. Our analysis in this section will consist of two parts. First, we investigate how our frames affect beliefs and contributions. We then turn to a test of guilt aversion and reciprocity (H_1 and H_2) as explanations for the observed contribution behavior. To make the data analysis between our TAKE and GIVE treatments comparable we express everything in the size of the public good (*i.e.*, what people contribute to the public good in the GIVE treatments, or what subjects leave in the public good in the TAKE treatments).

Result 1 concerns how our frames have affected beliefs.

Result 1: *The frames strongly affected first- and second-order beliefs.*

Support: Figures 1 and 2 provide the main support for Result 1. Figure 1 shows the mean first-order beliefs (*i.e.*, $(b_{ij}+ b_{ik})/2$) and the confidence bounds.

Figure 1: *Mean first-order beliefs* (*and confidence intervals*) *for each treatment*

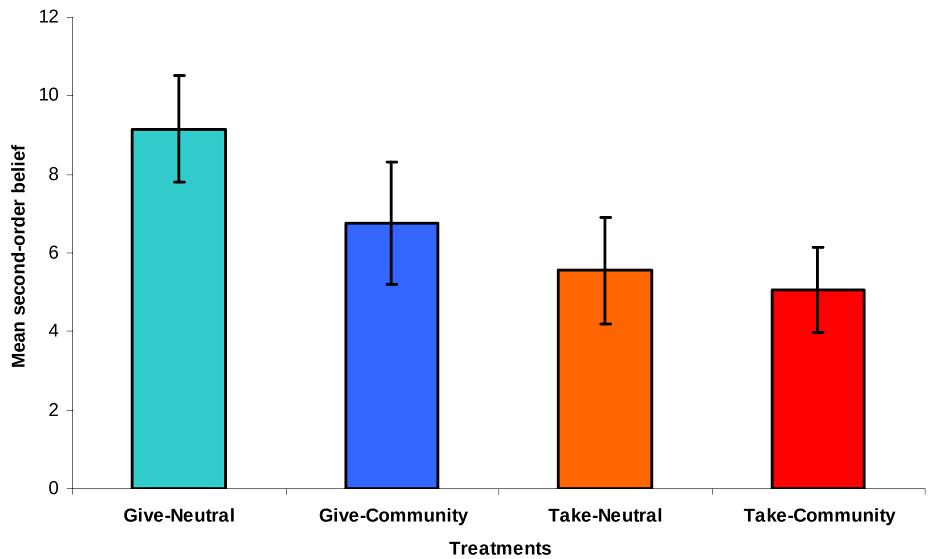


We find that first-order beliefs vary between 8 tokens in GIVE-NEUTRAL and slightly more than 4 tokens in TAKE-COMMUNITY. A non-parametric Kruskal-Wallis test strongly rejects the null hypothesis that the first-order beliefs from our four treatments stem from the same distribution ($\chi^2(3) = 18.28$, p=0.0004). A parametric analysis of variance (ANOVA) comes to the same conclusion (F=6.63; p=0.0003). The analysis of variance with respect to the factors "valence" and "label" shows that the factor "valence" is highly significant (F=15.81; p=0.0001); "label" is marginally significant (F=3.26; p=0.0722). The interaction variable "label×valence" is insignificant (F=0.00, p=0.9783). We conclude that both the context and in particular the valence framing affect the first-order beliefs. Most importantly, subjects in the TAKE treatments hold lower beliefs that others contribute than subjects in the GIVE treatments.

Figure 2 depicts the means and confidence bounds of the *second-order beliefs* (*i.e.*, $(c_{iji}+c_{iki})/2$). We find that the distributions of second-order beliefs are as well strongly and highly significantly affected by the frames (Kruskal-Wallis test, $\chi^2(3) = 21.97$, p = 0.0001). Again, an ANOVA supports this finding (F=8.11; p=0.0000): the factor "valence" is highly significant

(F=16.00; p=0.0001); "label" is significant at the five-percent level (F=4.75; p=0.0303); and the interaction variable "label×valence" is insignificant (F=2.08; p=0.1506). In other words, subjects in the TAKE treatments believe that the other group members expect them to contribute less than subjects in the GIVE experiments. Similarly, subjects believe that others expect them to contribute less in the COMMUNITY treatments than in the NEUTRAL treatments.

Figure 2: *Mean second-order beliefs* (*and confidence intervals*) *for each treatment*

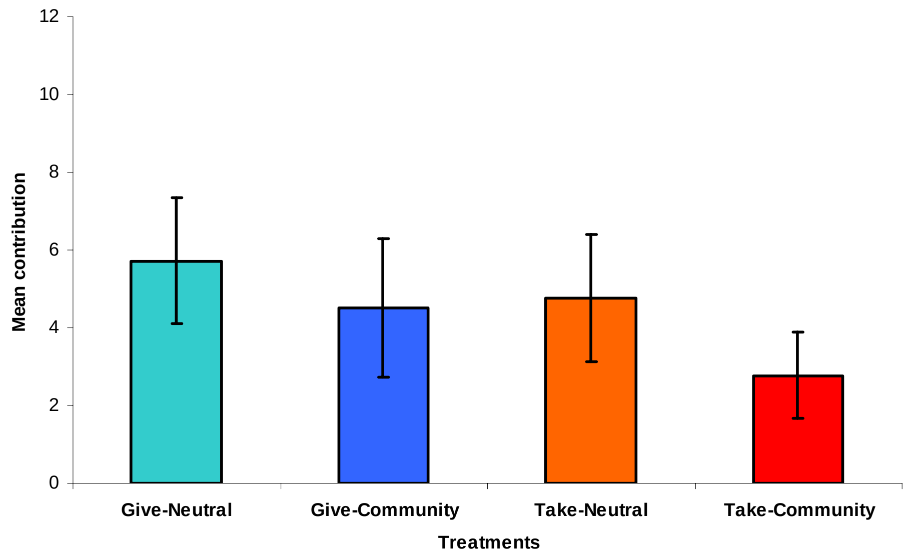


We now turn our attention to contributions. Before we investigate how beliefs have affected contribution behavior, we first look at contributions under the different frames. Result 2 records our findings.

Result 2: *The frames affected contributions but less strongly than beliefs.*

Support: Figure 3 provides the support for Result 2. Mean contributions are highest under GIVE-NEUTRAL and lowest under TAKE-COMMUNITY. A Kruskal-Wallis test suggests weakly significant differences between treatments ($\chi^2(3)=6.66$, p=0.0837). An ANOVA shows weakly significant treatment differences (F=2.63; p=0.0508): the factor "label" is significant (F=4.23; p=0.0408); "valence" is marginally significant (F=2.99; p=0.0850); and the interaction variable "label×valence" is insignificant (F=0.26; p=0.6119).[16,17]

Figure 3: *Mean contributions* (*and confidence intervals*) *for each treatment*



We will now turn our attention to the behavioral link between beliefs – which we have shown to be strongly affected by the frames – and behavior. Specifically, we will test our hypotheses H_1 and H_2, on guilt aversion and reciprocity.

Result 3 records the result concerning guilt aversion.

[16] If we drop this insignificant interaction variable, the model becomes significant (F=3.83; p=0.0231).

[17] A possibly surprising finding is that the community frame has lowered beliefs and contributions relative to the neutral frame. This result is in contrast to Ross and Ward (1996), Liberman et al. (2004) and Rege and Telle (2004), who are closest to our design. We have no explanation for this finding, except that it points at possible subject pool effects. Which beliefs subjects hold is an entirely empirical question. Frames may cue different beliefs in different subject pools because subject pools differ in background experiences.

Result 3: *The data support the guilt aversion hypothesis* H_1.

Support: Figure 4 and Table 2 contain the evidence in favor of Result 3. Figure 4 provides a graphical illustration of the guilt aversion hypothesis. In this figure we depict contributions as a function of the second-order beliefs (*i.e.*, $(c_{iji}+ c_{iki})/2$). The symbols represent combinations of contributions and second-order beliefs per treatment. The size of symbols is proportional to the underlying number of observations. Our hypothesis is that for subjects that contribute non-zero amounts contributions and second-order beliefs are positively correlated. We therefore distinguish in Figure 4 between zero contributions for positive second-order beliefs (indicated as filled circles on the *x*-axes) and the other contributions (indicated as triangles). The bold line is the trend line of the relationship between contributions and second-order beliefs (excluding the observations a_i=0&$(c_{iji}+ c_{iki})/2>0$).

Figure 4: *Contributions and second-order beliefs for each treatment*

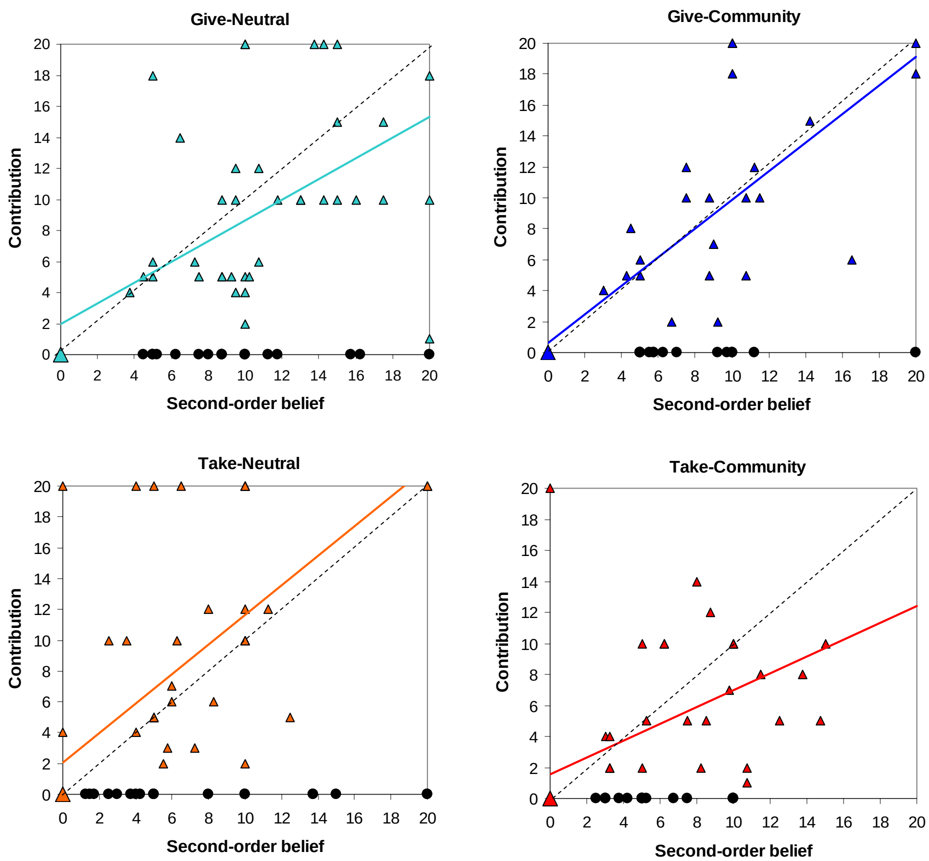


Figure 4 shows, first, that many subjects chose zero contributions even if they reported positive second-order beliefs: The fraction of subjects with positive second-order beliefs and zero contributions is very similar across all treatments ($\chi^2(3)=0.29$, $p=0.96$) and ranges from 27.8 percent in Take-Neutral to 31.8 percent in Take-Community. Second, contributions and second-order beliefs of subjects other than those who have a zero contribution despite a positive second-order belief are positively correlated in all four treatments. Table 3 corroborates this finding econometrically. Contributions and second-order beliefs are highly significantly positively correlated (t-values > 4.5).[18] Yet, the explained variance differs between treatments as

[18] This is actually true whether or not we include those subjects i for which $a_i = 0$ & $(c_{iji}+c_{iki})/2 > 0$.

the figure shows and as the regressions show formally. In GIVE-COMMUNITY, for instance, R^2 = 0.66, whereas in TAKE-COMMUNITY $R^2 = 0.31$.

Table 2: *Testing the guilt aversion hypothesis*

	Dependent variable: Contributions			
	GIVE-NEUTRAL	GIVE-COMMUNITY	TAKE-NEUTRAL	TAKE-COMMUNITY
Second-order beliefs	0.669	0.922	0.959	0.544
	(0.130)***	(0.105)***	(0.118)***	(0.119)***
Constant	1.957	0.644	2.035	1.527
	(1.069)*	(0.404)	(0.896)**	(0.876)*
Observations	47	36	52	45
R-squared	0.38	0.66	0.45	0.31

Notes: 1. OLS-regression; Robust standard errors in parentheses.
2. * significant at 10%; ** significant at 5%; *** significant at 1%.
3. Zero contributions for positive second-order beliefs are excluded.

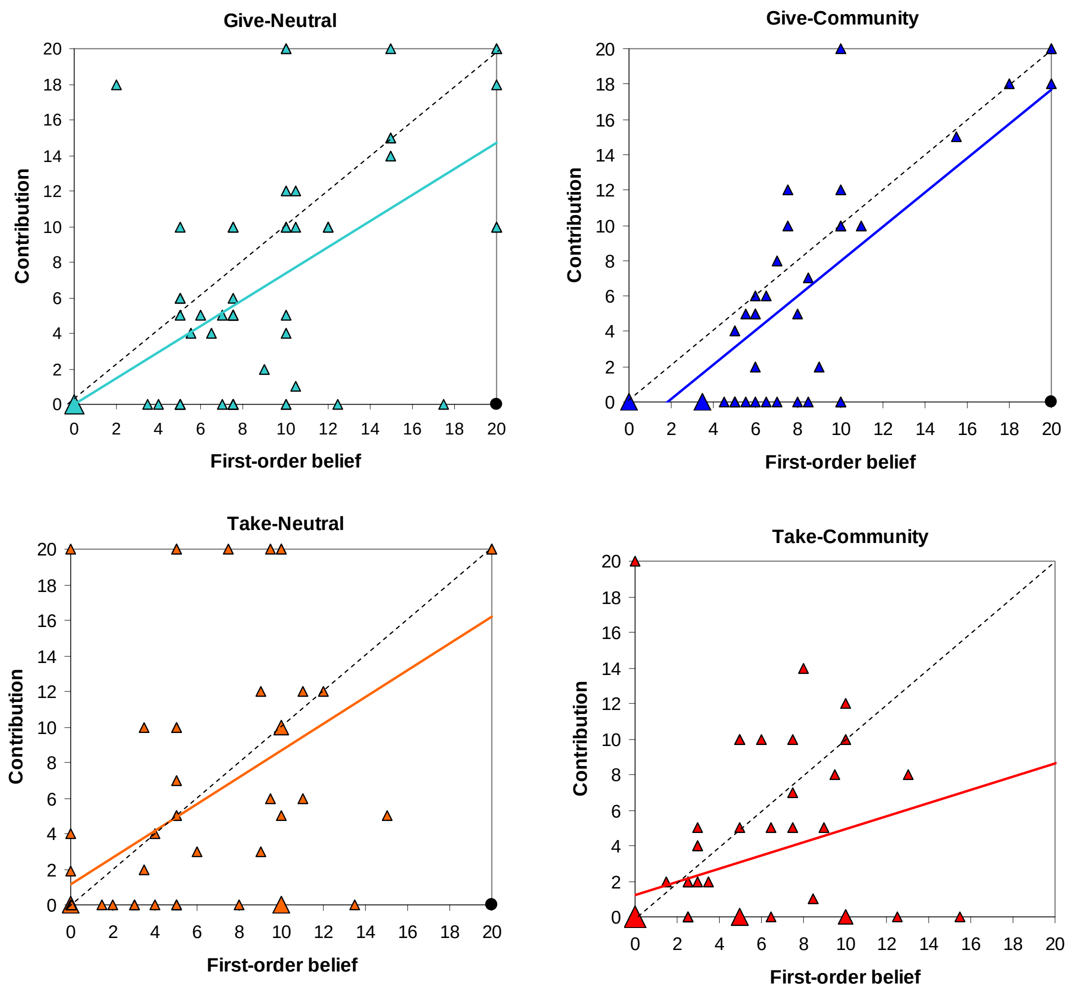
When we test whether the regression coefficients are significantly different across treatments, we find that the constants are the same across treatments ($F(3,172) = 1.08$; $p = 0.359$). The slopes, however, differ significantly across treatments ($F(3,172) = 2.94$, $p = 0.0349$), which implies that the frame affects the relationship between second-order beliefs and contributions. In other words, the frames also affect *how* guilt aversion shapes contribution behavior.

We turn now to reciprocity and H_2, which concerns the relation between a subject's contribution and his or her *first-order beliefs*. Figure 5 provides a graphical illustration. We distinguish between subjects who contribute nothing despite a first-order belief $(b_{ij}+ b_{ik})/2 = 20$; (indicated by filled circles on the *x*-axes), and the others (indicated by triangles). The size of symbols is proportional to the number of observations. The bold line is again the trend line (excluding observations $a_i=0\&(b_{ij}+ b_{ik})/2 = 20$).

Result 4: *The data support the reciprocity hypothesis H_2.*

Support: Figure 5 and Table 3 provide the support for Result 4.

Figure 5: *Contributions and first order beliefs for each treatment*



First, across all treatments, four participants contributed nothing despite holding an average first-order belief = 20. Second, contributions and first-order beliefs of the rest of the subjects are on average positively correlated. The trend line follows the diagonal quite closely in all treatments (except TAKE-COMMUNITY), which means that subjects matched their first-order belief on average, as predicted by our reciprocity hypothesis.

Table 3 corroborates these findings econometrically. We again confine our attention to subjects other than those who have a zero contribution despite an average first-order belief = 20. We find that all first-order belief coefficients are significantly positive (t-values vary between 14.4 in GIVE-COMMUNITY and 2.49 in TAKE-COMMUNITY).[19] Again, the explained variance differs between treatments. The R^2 is highest in GIVE-COMMUNITY (0.66) and lowest in TAKE-COMMUNITY (R^2 = 0.12). In other words, the link between first-order beliefs and contributions is tightest in GIVE-COMMUNITY and loosest in TAKE-COMMUNITY.

Table 3: *Testing the reciprocity hypothesis*

	Dependent variable: Contributions			
	GIVE-NEUTRAL	GIVE-COMMUNITY	TAKE-NEUTRAL	TAKE-COMMUNITY
First-order beliefs	0.735	0.973	0.752	0.371
	(0.115)***	(0.067)***	(0.128)***	(0.149)**
Constant	0.022	-1.771	1.187	1.23
	(0.895)	(0.503)***	(0.723)	(0.705)*
Observations	65	50	70	66
R-squared	0.40	0.66	0.31	0.12

Notes: 1. OLS-regression; Robust standard errors in parentheses.
2. * significant at 10%; ** significant at 5%; *** significant at 1%.
3. Zero contributions for average first-order beliefs = 20 are excluded.

When testing for the regression coefficients in Table 3 to be different from one another, we find the slopes to differ significantly between treatments (F(3,243) = 5.02; p = 0.0022). The constants are highly significantly different from one another (F(3,243) = 5.88; p = 0.0007). Thus, frames shift both the level and the slope of the relationship of first-order beliefs and contributions.

In summary, our results show that frames affect beliefs and beliefs affect contribution behavior. This finding can be embraced by psychological game theory, and indeed individual

[19] This is actually true whether or not we include those subjects *i* for which $a_i = 0$ & $(b_{ij}+b_{ik})/2 = 20$.

subject data on choices and beliefs exhibit support for psychological-game based theories of guilt aversion and reciprocity. This is the main finding of section 3.F.

4. CONCLUDING REMARKS

Framing effects are a challenge to traditional rational choice models, which assume description invariance. In this paper we first argued that psychological-game theoretic models can accommodate framing effects without reference to bounded rationality or cognitive biases. We then used public good experiments to examine the empirical relevance of this claim. We find support for two psychological-games based theories which can accommodate framing effects, namely guilt aversion and reciprocity.

There seems to be almost no prior theoretical work which attempts to explain framing effects. One shining exception is “variable frame theory” (VFT) (Bacharach 1993, Bacharach and Bernasconi 1997), which describes how players in a game conceptualize strategies and how this mental process affects play. This is different from our perspective; VFT deals with how players create frames while we look at how given frames affect choices. In section 3.D we explained how we propose to understand framing as a two-step chain: (i) frames move beliefs, and (ii) beliefs shape motivation and choice. If we were to factor in VFT we would have to add a step (o), which would precede step (i). It is thus meaningful to view VFT and our approach as complementary. We have made no attempt to link the approaches here, but future research may have such a goal in mind.

REFERENCES

Abbink, Klaus, and Heike Hennig-Schmidt (2006): Neutral versus Loaded Instructions in a Bribery Experiment, *Experimental Economics* 9, 103-21.

Andreoni, James (1995): Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments, *Quarterly Journal of Economics* 110, 1-21.

Aumann, Robert (1990): Nash Equilibria Are not Self-Enforcing, in Jean Jaskold Gabszewics, Jean-Francois Richard, and Laurence A. Wolsey (eds.): *Economic Decision Making: Games, Econometrics and Optimization*, *Essays in Honor of Jacques Dreze*, Elsevier Science Publishers, Amsterdam, 201-206.

Bacharach, Michael (1993): Variable Universe Games, in K. Binmore, A. Kirman and P. Tani (eds.), *Frontiers of Game Theory*, MIT Press.

Bacharach, Michael and Michele Bernasconi (1997): The Variable Frame Theory of Focal Points: An Experimental Study, *Games and Economic Behavior* 19, 1-45.

Bacharach, Michael, Gerardo Guerra and Daniel Zizzo (2004): *The Self-Fulfilling Property of Trust: An Experimental Study,* Oxford University Department of Economics Discussion Paper 76, 2001/2004.

Battigalli, Pierpaolo & Martin Dufwenberg (2005), "Dynamic Psychological Games," mimeo.

Baumeister, Roy, Arlene Stillwell and Todd Heatherton (1994): Guilt: An Interpersonal Approach, *Psychological Bulletin* 115, 243-67.

Baumeister, Roy, Arlene Stillwell and Todd Heatherton (1995): Personal Narratives about Guilt: Role in Action Control and Interpersonal Relationships, *Basic and Applied Social Psychology* 17, 173-98.

Brandts, Jordi and Christiane Schwieren (2006): Frames and Games, Paper Presented at the "International Workshop on Testing Different Theories of Individual Behavior in Social Dilemmas", January 11-13, 2006, Max Planck Institute of Economics, Jena.

Burnham, Terence, Kevin McCabe, and Vernon L. Smith (2000): Friend- or Foe: Intentionality Priming in an Extensive Form Trust Game, *Journal of Economic Behavior and Organization* 43, 57-74.

Camerer, Colin F. (2003): *Behavioral Game Theory*, Princeton, Princeton University Press.

Carlsson, Hans & Eric van Damme (1993), "Global games and equilibrium selection", *Econometrica* 61, 989-1018.

Clark, Ken, Steve Kay and Martin Sefton (2001), When Are Nash Equilibria Self-Enforcing? An Experimental Analysis, *International Journal of Game Theory* 29, 495-515.

Charness, Gary and Martin Dufwenberg (2006): Promises and Partnership. Forthcoming in *Econometrica*.

Charness, Gary (2000): Self-Serving Cheap Talk and Credibility: A Test of Aumann's Conjecture, *Games and Economic Behavior* 33, 177-94.

Cookson R. (2000): Framing Effects in Public Goods Experiments, *Experimental Economics* 3, 55-79.

Crawford, Vincent (1995), "Adaptive Dynamics in Coordination Games", *Econometrica* 63, 103-43.

Croson, Rachel (2000): Thinking like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play. *Journal of Economic Behavior and Organization* 41, 299–314

Croson, Rachel (2002): Theories of Altruism and Reciprocity: Evidence from Linear Public Goods Games. Mimeo.

Dufwenberg, Martin (2002): Marital Investment, Time Consistency, and Emotions, *Journal of Economic Behavior and Organization* 48, 57-69.

Dufwenberg, Martin and Uri Gneezy (2000): Measuring Beliefs in an Experimental Lost Wallet Game, *Games and Economic Behavior* 30, 163-82.

Dufwenberg, Martin and Georg Kirchsteiger (2004): A Theory of Sequential Reciprocity, *Games and Economic Behavior* 47, 268-98.

Elliott, Catherine S., Donald M. Hayward and Sebastian Canon (1998): Institutional Framing: Some Experimental Evidence, *Journal of Economic Behavior and Organization* 35, 455-64.

Falk, Armin and Urs Fischbacher (2006): A Theory of Reciprocity, *Games and Economic Behavior* 54, 293-315.

Fehr, Ernst and Simon Gächter (2000): Fairness and Retaliation: The Economics of Reciprocity, *Journal of Economic Perspectives* 14, 159-81.

Fischbacher, Urs (1999): z-Tree, Toolbox for Readymade Economic Experiments: Experimenter's manual. Working paper No. 21, Institute for Empirical Research in Economics, University of Zürich.

Fischbacher, Urs and Simon Gächter (2006): Heterogeneous social preferences and the dynamics of free riding in public goods. CeDEx Working Paper No. 2006-01, University of Nottingham.

Forsythe, Robert, Joel L. Horowitz, N.E. Savin and Martin Sefton (1994): Fairness in Simple Bargaining Experiments, *Games and Economic Behavior* 6, 347-69.

Geanakoplos, John, David Pearce and Ennio Stacchetti (1989): Psychological Games and Sequential Rationality, *Games and Economic Behavior* 1, 60–79.

Harsanyi, John and Reinhard Selten (1988), *A General Theory of Equilibrium Selection in Games*, MIT Press.

Huang, Peter and Ho-Mou Wu (1994): More Order without More Law: A Theory of Social Norms and Organizational Cultures, *Journal of Law, Economics and Organization* 10, 390-406.

Ledyard, John O. (1995): Public Goods: A Survey of Experimental Research, in: Kagel, John H. and Alvin E. Roth (eds.), *The Handbook of Experimental Economics*, Princeton University Press, Princeton, N.J, 253-348.

Levin, Irwin P., Sandra L. Schneider, and Gary J. Gaeth, (1998): All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects, *Organizational Behavior and Human Decision Processes* 76, 149-88.

Liberman, Varda, Steven Samuels and Lee Ross (2004): The Name of the Game: Predictive Power of Reputations vs. Situational Labels in Determining Prisoner's Dilemma Game Moves, *Personality and Social Psychology Bulletin* 30, 1175-85.

Mehta, Judith, Chris Starmer and Robert Sugden (1994): The nature of salience: an experimental investigation. *American Economic Review* 84, 658-73.

Ostrom, Elinor (2006): Framing a 2-Person Investment Game as a Common Pool, Paper Presented at the "International Workshop on Testing Different Theories of Individual Behavior in Social Dilemmas", January 11-13, 2006, Max Planck Institute of Economics, Jena.

Park, Eun-Soo (2000): Warm-Glow versus Cold-Prickle: A Further Experimental Study of Framing Effects on Free-Riding, *Journal of Economic Behavior and Organization* 43, 405-21.

Pruitt, Dean G. (1967): Reward Structure of Cooperation: The Decomposed Prisoners' Dilemma Game. *Journal of Personality and Social Psychology* 7, 21-27.

Rabin, Matthew (1993): Incorporating Fairness into Game Theory and Economics, *American Economic Review* 83, 1281-1302.

Rege, Mari and Kjetil Telle (2004): The Impact of Social Approval and Framing on Cooperation in Public Goods Situations, *Journal of Public Economics* 88, 1625-44.

Ross, Lee, and Andrew Ward (1996): Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding, in: Reed, Edward S., Elliott Turiel and Terrence Brown (eds.), *Values and Knowledge.* Lawrence Erlbaum Associates, Mahwah, N.J., 103-35.

Schelling, Thomas C. (1960): *The Strategy of Conflict*, Harvard University Press, Cambridge, MA.

Schwartz-Shea, Peregrine, and Randy T. Simmons (1995): Social Dilemmas and Perceptions: Experiments on Framing and Inconsequentiality, in: Schroeder, David A. (ed.), *Social Dilemmas. Perspectives on Individuals and Groups.* Praeger Publishers, Westport, CT.

Selten, Reinhard, and Claus C. Berg (1970): Drei experimentelle Oligopolserien mit kontinuierlichem Zeitverlauf, in: Sauermann, Heinz (ed.), *Beiträge zur experimentellen Wirtschaftsforschung*, Vol. 2, J.C.B. Mohr, Tübingen, 162-221.

Sobel, Joel (2005): Interdependent Preferences and Reciprocity, *Journal of Economic Literature* 43, 392-436.

Sonnemans, Joep, Arthur Schram, and Theo Offerman (1998): Public Good Provision and Public Bad Prevention: The Effect of Framing, *Journal of Economic Behavior and Organization* 34, 143-61.

Tversky, Amos, and Daniel Kahneman (1981): The Framing of Decisions and the Psychology of Choice, *Science* 211, 453-58.

Willinger, Marc, and Anthony Ziegelmeyer (1999): Framing and Cooperation in Public Good Games: An Experiment with an Interior Solution, *Economics Letters* 65, 323-28.

Appendix

Note: Text in brackets denotes the [GIVE treatments].

Instructions to the (community) experiment

Welcome to the *(community)* experiment

General information on the *(community)* experiment

You are now participating in an economic experiment which is financed by the European Union. If you read the following explanations carefully, you'll be able to earn a considerable amount of money – depending on your decisions. Therefore it is important to actually read the instructions very carefully.

The instructions are for your private information only**. During the experiment it is not allowed to communicate with other participants in any way.** If you have questions, please consult us.

During the experiment, we will not talk about Euro, but about Taler. Your total income will first be calculated in Taler. The total amount of Taler that you have accumulated during the experiment will be converted into Euro at the end of the experiment at an exchange rate of

1 Taler = 0.50 Euros

At the end of the experiment, you will be paid the total amount of Taler earned during the experiment and converted into Euro **in cash**.

At the beginning of the experiment, all participants will be randomly divided into groups of three. Besides you, there will be two more members in your group. **You will neither learn before nor after the experiment, who the other persons in your group are**.

The experiment consists of only one task. You have to decide **how many Taler you take from [contribute to] a project of your group and how many Taler you leave in the project [keep for yourself].** On the following pages we will describe the exact course of the experiment. At the end of this introductory information we ask you to do several control exercises which are designed to familiarize you with the decision situation.

The decision in the (*community*) experiment

At the beginning of the first stage, there are **60 Taler** in a project of your group [every participant receives an "endowment" of **20 Taler]**. You then have to decide how many of these 60 Taler you take from the project for yourself or how many you leave in the project. [You then have to decide how many of these 20 Taler you contribute to the project or how many you keep for yourself.] Each participant can take up to **20 Taler** from the project [can contribute up to 20 Taler]. The two other members of your group have to make the same decision. They can also either take Taler from the project for themselves or leave Taler in the project. [They can also either contribute Taler to the project or keep Taler for themselves.] You and the other members of the group can choose any amount to be taken [contribution] between 0 and 20 Taler.

Every Taler that you take from the project for yourself [do not contribute to the project] automatically belongs to you and will be paid to you, converted by the exchange rate given above, at the end of the experiment.

The following happens to the Taler that are not taken from [that are contributed to] the project: The project's value will be multiplied by 1.5 and this amount will be **divided equally among all three members of the group**. If for instance 1 Taler is not taken from [is contributed to] the project, the Taler's value increases to 1.5 Taler. This amount is divided equally among all three members of the group. Thus every group member receives 0.5 Taler.

Your income from the project rises by 0.5 Taler if you take one Taler less from [contribute one Taler more to] the project. At the same time, the income of the other two members of the group also rises by 0.5 Taler, because they receive the same income from the project as you do. Therefore, if you take one Taler less from [contribute one Taler more to] the project the income from the project with regard to the whole group increases by 1.5 Taler. It also holds that your income rises by 0.5 Taler if another group member takes one Taler less from [contributes one Taler more to] the project.

After all three members of the group have made their decisions about the amounts they take from [their contributions to] the project the total income achieved by each participant is determined.

How is your income calculated from your decision?

The income of every member of the group is calculated in the same way. The income consists of two parts:

(1) the Taler that somebody takes [keeps] for himself/herself **("income from Taler taken [kept]"**)
(2) the **"income from the project"**. The income from the project is
1.5 × (60 - sum of all Taler taken from the project)/3 =
0.5 × (60 - sum of all Taler taken from the project)
[1.5 × (sum of all Taler contributed to the project)/3 =
0.5 × (sum of all Taler contributed to the project)].

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

Your total income =

Income from Taler taken [kept] + Income from project =

(Taler taken by you) + 0.5 × (60 - sum of all Taler taken from project)

[(20 – Taler you contributed to project) + 0.5 × (sum of all Taler contributed to project)]

If you take all 20 Taler from [do not contribute anything to] the project, your "income from Taler taken [kept]" is 20. If you take [contribute] for instance 10 Taler from [to] the project, your "income from Taler taken [kept]" is 10. At the same time, the total sum of Taler left in [contributed to] the project decreases [increases] and so does your "income from the project".

In order to explain the income calculation we give some examples:

- If each of the three members of the group takes 20 Taler from [contributes 0 Taler to] the project, all three will receive an "income from Taler taken [kept]" of 20. Nobody receives anything from the project, because no one left [contributed] anything. Therefore, the total income of every member of the group is 20 Taler.
  *Calculation of the total income of every participant:(20) + 0.5 * (60-60) = 20*
  [*Calculation of the total income of every participant:(20 – 0) + 0.5 * (0) = 20*]
- If each of the three members of the group takes 0 [contributes 20] Taler there will a total of 60 Taler left in [contributed to] the project. The "income from Taler taken [kept]" is zero for everyone, but each member receives an income from the project of 0.5 * 60 = 30 Taler.
  *Calculation of the total income of every participant:(0) + 0.5 * (60-0) = 30*
  [*Calculation of the total income of every participant:(20 – 20) + 0.5 * (60) = 30*]
- If you take 0 [contribute 20] Taler, the second member 10 and the third member 20 [0] Taler, the following incomes are calculated.
  - Because the second and third member have together taken 30 Taler [you and the second member have together contributed 30 Taler], everyone will receive 0.5 * 30 = 15 Taler from the project.
  - You took 0 [contributed all your 20] Taler from [to] the project. You will therefore receive 15 Taler in total at the end of the experiment.
  - The second member of the group also receives 15 Taler from the project. In addition, he receives 10 Taler "income from Taler taken [kept]" because he took [contributed only] 10 Taler from [to] the project [Thus, 10 Taler remain for himself], and he receives 15 + 10 = 25 Taler altogether.
  - The third member of the group, who took all Taler [did not contribute anything], also receives the 15 Taler from the project and additionally the 20 Taler "income from Taler taken [kept]", which means 20 + 15 = 35 Taler altogether.

  *Calculation of your total income: (0) + 0.5 * (60-30) = 15*
  *Calculation of the total income of the 2nd group member: (10) + 0.5 * (60-30) = 25*
  *Calculation of the total income of the 3rd group member: (20) + 0.5 * (60-30) = 35*
  [*Calculation of your total income: (20 – 20) + 0.5 * (30) = 15*
  *Calculation of the total income of the 2nd group member: (20 – 10) + 0.5 * (30) = 25*
  *Calculation of the total income of the 3rd group member: (20 – 0) + 0.5 * (30) = 35*]
- The two other members of your group take 0 [contribute 20] Taler each from [to] the project. You take all Taler [do not contribute anything]. In this case the income will be calculated as follows:
  *Calculation of your total income (amount taken 20): (20) + 0.5 * (60-20) = 40*

Calculation of the total income of the 2^{nd} and 3^{rd} group member (amount taken 0):
*(0) + 0.5 * (60-20) = 20*
[*Calculation of your total income (contribution 0): (20 – 0) + 0.5 * (40) = 40*
Calculation of the total income of the 2^{nd} and 3^{rd} group member (contribution 20):
*(20 – 20) + 0.5 * (40) = 20*]

When making your decision you will see the following screen:

Please make the decision on the amount to be taken by you [your contribution] in the *(community)* experiment now.

In the project, there are [Your endowment] 60 [20]
The amount to be taken by you from [Your contribution to] the project…..

You will make your decision on a screen like the one above and enter into the blank space how many Taler you take from [contribute to] the project.

After you have made your decision please press the OK-button. As long as you did not press the button you can change your decision anytime.
The experiment will be carried out **once**.

First order belief statement (text of questions)
After you have taken your decision in the (*community*) experiment we would like to ask you for the following statement:

Please estimate how many Taler **the other two members of the group have taken from [contributed to] the project in total**.

If you estimated the correct amount you will be paid **20 EURO**.

Example 1:
You estimate that the other two members of the group took [contributed] 31 Taler from [to] the project. In fact, both members took [contributed] 19 and 12 Taler. Your estimation was correct and you will be paid **20 EURO**

Example 2:

You estimate that the other two members of the group took [contributed] 17 Taler from [to] the project. In fact, both members took [contributed] 12 and 6 Taler. Your estimation was wrong and you will be paid **0 EURO**

(Note that your estimation must be a number between 0 and 40 including these numbers.)

Estimated amount taken by [contribution of] the other two group members in the *(community)* experiment in total:

Second order belief statement (text of questions)

Each member in your group has estimated in the same way as you did how many Taler in total the other two members of the group took from [contributed to] the project.

Please estimate now the sum of amounts the other two group members **stated** as estimation in the (*community*) experiment.

If you estimated the correct amount you will be paid **20 EURO**.

Example 1:

You estimate that the other two members of the group stated an estimation of 57 Taler. In fact, the second member stated 31 and the third member stated 26 Taler as estimation.

Your estimation was correct and you will be paid **20 EURO**

Example 2:

You estimate that the other two members of the group stated an estimation of 42 Taler. In fact, the second member stated 17 and the third member stated 21 Taler as estimation.

Your estimation was wrong and you will be paid **0 EURO**

(Note that your estimation must be a number between 0 and 80 including these numbers.)

Estimated sum of amounts the other two group members stated as estimation in the *(community)* experiment: