# Is observed other-regarding behavior always genuine?

Astrid Matthey [♠*] Tobias Regner [♣]

[♠] *Max Planck Institute of Economics, Jena, Germany*

[♣] *Friedrich Schiller University, Jena, Germany*

September 18, 2008

**Abstract**

We investigate to what extent genuine social preferences can explain observed other-regarding behavior. In a social dilemma situation (a dictator game variant), subjects can choose whether to learn about the consequences of their choice for the receiver. We find that a majority of subjects showing other-regarding behavior when the payoffs of the receiver are known choose to ignore them if possible. This behavior is inconsistent with genuine social preferences.

Our model explains other-regarding behavior as avoiding cognitive dissonance: People do not behave fairly because they genuinely care for others, but because they like to think of themselves as being fair. The model can explain our data as well as earlier experimental data.

*JEL classifications*: C72, D01, C91, D80

*Keywords*: social preferences, experiments, social dilemma, cognitive dissonance

[*]Corresponding address: matthey@econ.mpg.de, phone: +49 3641 686644.

We would like to thank seminar audiences at the Trento workshop on Behavioral Economics, the EEA/ESEM 2008 joint congress in Milano, the IAREP/SABE 2008 congress in Roma, the Max Planck Institute of Economics in Jena and the Comenius University Bratislava. We are grateful to Rosemarie Nagel and Ondřej Rydval for valuable comments.

# 1 Introduction

By now a substantial body of literature exists on other-regarding behavior. While the empirical evidence is abundant and unanimously confirming other-regarding behavior, the theoretical side remains less unified.[1] The aim of this paper is to further contribute to the research on social preferences. In particular, we try to shed some more light on the actual process that leads to behavior that is perceived as "other-regarding".

Participants in our experiment face some standard "full information" social dilemma situations, that is, situations where they are fully informed about the payoff of the participant affected by their choice. In addition, they face social dilemma situations where they have the possibility to remain uninformed about the consequences on the other participant's payoff: People know that a social dilemma exists, i.e., that their behavior affects others. But they can ignore the exact consequences of this behavior. They can - but do not have to - find out which of several possible effects their behavior has on others.

Previous studies[2] analyzed the effect of varying levels of information in social dilemma situations when subjects can avoid entering the social dilemma situation at all. Our design differs from this. We are rather interested in whether there is a change in participants' behavior given that they remain in a social dilemma: Do people who show other-regarding behavior under "full information" choose to find out the consequences of their choice if they can stay uninformed? In other words, are they "genuine pro-socials"? Or do they act pro-socially under full information, yet prefer to remain ignorant when given the chance (taking negative consequences for others into account)? Furthermore - if we find this behavior - we are also interested in the driving motivational forces that would distinguish such "ignorers" from "genuine pro-socials".

In order to study such ignoring behavior we develop a model that is based on cognitive dissonance, a psychological theory introduced by Festinger (1957).

[1] Two approaches can be distinguished, outcome-based and intentions-based models. Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) are seminal papers of the first strand, Rabin (1993), Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2005) of the second. Both approaches have their respective advantages and drawbacks as illustrated in surveys on the literature (Camerer (2003) or Fehr and Schmidt (2003)).

[2] Dana et al. (2006), Lazear et al. (2006) and Broberg et al. (2007).

A person experiences cognitive dissonance when she holds two psychologically conflicting cognitions, say, she finds a certain task boring, but claims that it was interesting as an internal justification for actually doing it (Festinger 1957). We use this early, general theory rather than more recent but narrower ones like, e.g., self-affirmation theory (Steele 1988), since it provides sufficient structure to explain observed behavior without unduly restricting the range of possible influences.

In our experiment cognitive dissonance may affect participants' behavior. If they prefer a fair split under full information but choose to ignore if given the chance, they will have two inconsistent psychological cognitions: in one situation they behave fairly, in the other one they do not. Hence, they should experience cognitive dissonance, a negative psychological cost. Ignoring pro-socials bear this cost in order to gain a higher material payoff, while genuine pro-socials avoid the potential psychological conflict and choose a fair outcome.

The results of the experiment show that indeed a significant number of subjects showing other-regarding behavior under full information prefer to remain ignorant if this is possible. Our cognitive dissonance-based model can explain this split into ignoring and genuine pro-socials, while other social preferences models cannot. In addition, our experimental design enables us to test aspects of the suggested model. The study of Dana et al. (2007) supports our experimental results. However, they use a setup that does not allow the test of a theory. Ours does.

The experimental design reflects features of many decisions we make every day. We often know that a choice involves a social dilemma, but can remain uninformed about the exact consequences of this choice. Whenever we shop for food, clothes, etc., we face the decision of whether to pay higher prices for goods that were produced under higher social and/or ecological standards. This means that with each purchase decision we choose whether to share our money with others, who then get higher wages, better social insurance, health care etc. Before we make this decision, however, we decide whether to pay attention to the way of production, or to whether a product is traded fairly. We can avoid the sharing decision if we simply pick the product which we like best, and ignore the way it was produced and traded.

These are the characteristics we simulate in our experiment. In contrast, the sorting experiments of Dana et al. (2006) and Lazear et al. (2006) offer

people an alternative where their behavior does not have a direct impact on others anymore. This alternative may be available for the beggar on the street, whom we can avoid by crossing the road. It is rarely available when we purchase food, clothes or other consumption goods.

The paper is organized as follows. In section 2 we describe the experimental design. Section 3 develops the model and applies it to our experimental setup. Results are presented in section 4. Section 5 concludes.

# 2 Experiment

## 2.1 Design

The experiment consisted of four periods. In periods 1 to 3, each subject chose among five options, A to E. Each option assigned one outcome to the subject herself, and one to another person, named participant Y. A third participant, called participant X, also assigned one outcome to himself and one to the subject. This means that in each period, each subject received one outcome from a participant called X, and chose an option which assigned one outcome to herself, and one to another participant called Y. Participants X and Y were chosen randomly, and unknown to the subject. Matching of subjects took place according to a perfect stranger design.

In each period, the five outcomes assigned by the options were 1 Euro, 1 Euro, 4 Euro, 4 Euro and 7 Euro, for all participants. This was common knowledge. Each option a subject could choose assigned one of these outcomes to the subject herself, and one to the subject in the role of participant Y. How the two sets of outcomes were combined was determined randomly. For example, an outcome of 7 Euro for the subject herself could induce an outcome of 7 Euro for Y (with 20% chance), of 4 Euro (40% chance) or of 1 Euro (40% chance). These probabilities were also stated in the instructions.

In periods one and three, subjects could initially only see their own outcomes, but not the outcomes their choice would assign to Y. Subjects could then decide whether to i) choose an option immediately, i.e., without knowing the consequences for Y, ii) uncover the outcomes that each option assigned to participant Y, and then choose an option, or iii) see the outcome pairs from which participant X could choose, and then choose an option (their own, not that of X). Subjects were aware that knowing the options of

X would not allow them to influence his choice.
In period two, the outcomes of the subject herself and participant Y were visible immediately, and subjects directly chose an option. In fact, the options (outcome pairs) in periods one and two were the same. Only their order was reversed to make this fact less obvious.

In period four subjects faced a series of options. First they had to choose between an outcome of 10 Euro for themselves and 0 for participant Y, versus equal outcomes for both. The equal outcomes increased from 0 Euro for both to 10 Euro for both in steps of 50 Cents.

Second, subjects had to choose between the following pairs of outcomes, where the first number of each pair denotes the subject's own outcome (in Euro) and the second the outcome of participant Y: (4,4) vs. (4,7); (4,1) vs. (4,7); (7,1) vs. (4,4); (7,1) vs. (4,7). Only one of the 25 choices in period four was possibly paid. We used the strategy method in order to get a more comprehensive picture of the subjects' preferences.

After all choices were made, one period was chosen randomly for each individual. From this period, the individual received either the outcome she assigned to herself, or the outcome that participant X assigned to her. If period four was drawn, the decision to be paid was determined randomly in a second step.

After the experiment was completed, subjects filled in a questionnaire. Among other questions, they were asked how hard it was for them to decide whether they wanted information on the outcomes of participant Y, or on the options of participant X, and whether they found it easy to choose an option. Further, they were asked for the share of subjects they expected to have uncovered the outcomes of participant Y in periods one and three, how much they thought "their" participant Y received if they did not uncover his outcomes, and what they thought appropriate behavior would have been in the respective periods.

## 2.2 Participants and Procedures

The experiment took place at the laboratory of the Max Planck Institute of Economics in Jena, Germany. 99 participants were recruited among students from various disciplines at the University of Jena. In each session gender composition was approximately balanced and subjects took part only in one

session. Participants were recruited using the ORSEE software (Greiner, 2004). The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007) and took, on average, 50 minutes. The average earnings in the experiment have been of €7.61. The show-up fee for the experiment was €2.5.

Upon arrival at the laboratory subjects were randomly assigned to one of the computer terminals. Each computer terminal is in a cubicle that does not allow communication or visual interaction among the participants. Participants were given time to privately read the instructions. Subjects were allowed to privately ask clarifications about the instructions. To check the understanding of the instructions subjects were asked to answer some control questions. After all subjects had answered the questions correctly the experiment started.

At the end of the experiment subjects were paid in cash according to their performance. Privacy was warranted during the payment phase.

# 3 Theory

## 3.1 Other-regarding preferences and cognitive dissonance

Observed other-regarding behavior is usually explained with other-regarding preferences. This means that people are assumed to be truly concerned about how much others have. For example, in the Fehr-Schmidt model (Fehr and Schmidt, 1999), people's preferences are assumed to be of the (simplified) form

$$U(x_i, x_{-i}) = u(x_i) - \alpha \max\{x_{-i} - x_i, 0\} - \beta \max\{x_i - x_{-i}, 0\} \qquad (1)$$

where $x_i$ is the individual's own payoff, $x_{-i}$ is the other person's payoff, $\alpha$ is the individual's aversion to unfavorable inequity and $\beta$ is her aversion to favorable inequity.

This means that the individual truly values equal payoffs.

Sorting Experiments (Lazear et al., 2006, Dana et al., 2006) have shown that many individuals do not really value equal payoffs, although their behavior may reflect sharing if put into social dilemma situations. If given the choice, many subjects in these experiments avoided the social dilemma situation, even if this led to reductions in their maximum payoff.

Lazear et al. (2006) formalized these results by assuming that some people "like sharing", others "like not-sharing", and quite a high number "dislike not-sharing". The latter group share their payoffs if put into a social dilemma situation ("asked to share"), but avoid this situation if possible.

Although such "like sharing", "like not-sharing" and "dislike not-sharing" preferences can explain the behavior observed in the sorting experiments, they are hard to generalize, and are somewhat disconnected from the psychological evidence.[3]

In this paper we choose a different approach, and attempt to model "other-regarding preferences" as based on psychological evidence. This attempt is motivated by the observation that other-regarding behavior to a considerable part seems not to be due to genuine other-regarding preferences. This is true even if the setup is anonymous and one-shot, i.e., group pressure, reputation etc. cannot explain the behavior.

The concept we employ to explain such behavior is not new. It was first described as "cognitive dissonance" by Leon Festinger (1957) as the negative drive state that arises if a person holds two cognitions that are psychologically inconsistent. For example, there may be a dissonance between a person's beliefs and her behavior, which is experienced as unpleasant, and produces a motivation to reduce this dissonance.

In social dilemma situations, people may have a certain perception about what they should do, what they would like themselves to do, what the person they would like to be would do etc. This is where preferences come in.

Rather than having genuine preferences regarding the payoffs of others, people may have a preference over the existence of a certain moral or ethical standard in society, and of their role in forming it, or contribution to it. For example, they may have a preference for a world where fair behavior is the standard, and acknowledge that if they want such a world, they should behave fairly too. Alternatively, some people may have a preference for a world where it is the agreed standard that everyone behaves in a rational and selfish way, and no one is blamed for that. From their preference for an ethical standard people then derive the picture of the person as that they see themselves, that they would like to be etc.

In this sense, people may have a preference regarding "behavior towards

[3]Due, presumably, to the different focus of the authors.

others", and hence also regarding *their own* behavior towards others, but not necessarily regarding these others directly. In the examples above, one may have a preference for people treating each other fairly, but beyond that not care much about the outcomes. In other words, people may have a preference for a *standard of good behavior*, rather than for particular outcomes.

If people behave in a way that deviates from the standard by which they measure themselves, they experience cognitive dissonance (e.g., Aronson, 1997). Since this results in unpleasant feelings, they like to avoid such deviations. As we will show below, this reasoning can explain standard other-regarding behavior, but also the behavior we observe in the experiment, which is not consistent with earlier models.

### 3.2 Cognitive dissonance in social dilemma situations

First we would like to emphasize that this is not meant as a general model of cognitive dissonance, based on all the available psychological evidence. Rather, it is an attempt to explain certain patterns of behavior observed in social dilemma situations based on the concept of cognitive dissonance.

Let $X$ denote the choice set of the individual. Elements of $X$ are $(x_i, x_{-i})$, with $x_i$ as the outcome that the individual's choice assigns to herself, and $x_{-i}$ as the outcome this choice assigns to another person. More generally, elements of $X$ could be defined as $(x_i, x_{-i1}, .., x_{-in})$, with $n$ other people being affected by the choice of individual $i$. For simplicity, and to reflect our experimental setup, in what follows we will restrict choices to affect only one other person, i.e., $n = 1$.
Let further $\mathcal{X}$ denote the set of choices the individual expects relevant others would make if faced with the choice set $X$. Elements of $\mathcal{X}$ are $(x_j, x_{-j})$, with $j = 1, .., m$ as the group of relevant others that the individual compares herself to, and $(x_j, x_{-j})$ as the outcome that individual $i$ expects individual $j$ to assign to herself and another person. Since the choices considered in $\mathcal{X}$ are based on the available choices in $X$, all elements of $\mathcal{X}$ are also elements of $X$: $\mathcal{X} \subseteq X$. Note that for the model, it is unimportant how the "reference group" is formed, that is, whose choice the individual includes in $\mathcal{X}$.

Individual behavior can be classified into certain standards. Let $\Delta$ denote the set of behavioral standards or rules. An element $\Delta$ of $\Delta$ then denotes the behavioral standard of "fairness", "cleverness", "generosity" etc. Each

standard can occur in various degrees, e.g., $\Delta_1$, $\Delta_2$ and $\Delta_3$ may denote "very unfair", "slightly unfair" and "fair" behavior. How many elements $\mathit{\Delta}$ contains depends on the individual's ability (or willingness) to distinguish between behavioral rules.

Each individual has a preference relation $\succeq$ over the set of behavioral standards, where $\succ$ denotes strict preference and $\sim$ denotes indifference. For example, one individual prefers fair behavior, while another may prefer smart behavior or generous behavior. Individual preference relations are assumed to be complete and transitive.[4]

We then define a function $f$ which maps behavioral standards into choices, dependent on the available choice set and the individual's expectation of others' choices: $f(\Delta, X, \mathcal{X}) : \mathit{\Delta} \times X \times X \rightarrow X$. It should be noted that the function may also depend, e.g., on the individual's perception of what the individual who is affected by her choice expects her to choose (see, e.g., Dana et al., 2007). However, since there is little evidence on this factor, and we cannot test it in our experiment, we do not include it here. For a model including this factor, see Matthey and Regner (2007).

If $X$ and $\mathcal{X}$ are known, by inverting $f$ standards of behavior can be inferred from observed behavior. However, if the choice set is restricted, $f$ may assign the same choice to several behavioral standards. In this case, we will assume that the standard implied by observed behavior is the one that among all the standards consistent with the behavior is preferred by the individual. For example, the standards "slightly unfair" and "fair" may induce the same choice if only few choices are available. For an individual that prefers fairness, if she makes this choice we will judge her behavior as fair. Given this assumption, with a slight abuse of notation we can write $f^{-1}(x_i, x_{-i})$, and say that $f^{-1}$ assigns a unique (preferred) standard $\Delta$ to any choice $(x_i, x_{-i})$.

The set of available choices $X$ is included in $f$ to account for the effect of choice restrictions on the evaluation of behavior. If, for example, in a dictator game sharing payoffs 50-50 is not possible, but only sharing 30-70 or 70-30, sharing 70-30 will be evaluated differently than if the equal split is available. Similarly, if a split of 60-40 is the result of a real effort task,

[4]This may be a tough assumption when applied to different contexts and long time intervals. When applied within the context of a particular situation with limited duration, however, it seems an acceptable restriction.

keeping 60% of the payoff will be evaluated differently than in a normal dictator game. $\mathcal{X}$ is included to reflect the effect of social comparison that has been found in many areas of individual decision making or utility.

The utility that the individual derives from her decision is then defined as

$$\begin{aligned} U(x_i, x_{-i}) &= u(x_i) - D(\Delta^*, \Delta) \\ &= u(x_i) - D(\Delta^*, f^{-1}(x_i, x_{-i})) \quad . \end{aligned} \tag{2}$$

$u(x_i)$ is the utility that the individual derives from the outcome she herself receives as the consequence of her decision. The general form of this utility function is similar to the model of Rabin (1994). In contrast to his model, however, we analyze cognitive dissonance in detail. $D(\Delta^*, \Delta)$ is the cognitive dissonance the individual experiences from her choice. This dissonance depends on how the behavioral standard $\Delta$ that is consistent with her behavior compares to her desired behavior $\Delta^*$:[5]

$$\begin{aligned} &\text{If} \quad f^{-1}(x_i, x_{-i}) = \Delta \prec \Delta^* \qquad \Rightarrow \qquad D(\Delta^*, \Delta) > 0; \\ &\text{If} \quad f^{-1}(x_i, x_{-i}) = \Delta \succeq \Delta^* \qquad \Rightarrow \qquad D(\Delta^*, \Delta) = 0; \end{aligned}$$

In a standard dictator game, for example, increasing $x_i$ above 50% increases the individual's $u(x_i)$. However, if the behavior $\Delta^*$ she finds desirable is to behave fairly, and she finds sending less than 50% to be consistent only with unfair behavior, keeping more than half of the endowment leads to cognitive dissonance: $D(\Delta^*, f^{-1}(x_i > 50\%, x_{-i} < 50\%)) > 0$.
We will assume that in the short run $\Delta^*$ is constant: people do not change quickly the behavioral standard they consider as moral or appropriate. Hence, where no confusion can arise we will drop $\Delta^*$ in (2) and write $U(x_i, x_{-i}) = u(x_i) - D(f^{-1}(x_i, x_{-i}))$.

### 3.3 Cognitive Dissonance in the experiment

In the experiment we test whether people that show other-regarding behavior when faced with a social dilemma situation directly, i.e., when knowing the exact consequences of their behavior for others, choose to ignore these consequences if given the opportunity. Such behavior cannot be explained by genuine other-regarding preferences: People with a true preference for

[5] This ignores possible pride from behavior which is consistent with a standard higher than $\Delta^*$., i.e., $D < 0$.

sharing would like to make sure that they really choose the action that gives the highest utility. Hence, they would want to learn the consequences of their decision for the other person.

The hypothesis we want to test is that differences in cognitive dissonance drive people's behavior. In particular, we suggest that the motive for people to share their payoffs is to avoid cognitive dissonance. If dissonance can be avoided or sufficiently reduced while keeping high payoffs for oneself, the same people who shared before choose to ignore the consequences of their behavior.

The situations in the experiment can be formalized as follows. Consider first the case when the individual can freely observe all payoffs, i.e., she knows the exact consequences of her behavior (period 2 of the experiment). The functions which apply to these situations receive the index $F$ for "full" information. Choice $(x_i, x_{-i})$ given choice set $X_F$ then induces utility

$$U_F(x_i, x_{-i}) = u(x_i) - D(f_F^{-1}(x_i, x_{-i})) \quad .$$

The function $f_F$ maps standards to choices, given the choice set $X_F$ that assigns an $x_{-i}$ to each $x_i$ the individual may choose. Given this choice set, the individual forms $\mathcal{X}_F$.

Consider now the case where initially the individual does not know the exact consequences of her decision, but has to decide whether to learn or ignore them (periods 1 and 3 of the experiment). Her utility in case of ignorance (index $I$), i.e., choosing an alternative without learning about the consequences, is then given as

$$U_I(x_i, x_{-i}) = u(x_i) - D(f_I^{-1}(x_i, \hat{x}_{-i})) \quad . \tag{3}$$

For example, she can take the maximum payoff for herself, and experiences dissonance depending on the payoff she believes the other will get given her choice, $\hat{x}_{-i}$, and knowing that she could have uncovered the payoffs. The latter is relevant since feelings of dissonance can be expected to depend on whether the ignorance was voluntary or enforced. Accordingly, the function $f_I$ maps standards to choices given the choice set $X_I$ and voluntary ignorance. Again, given $X_I$ the individual will form a set of expectations $\mathcal{X}_I$ over the choices of others.

For rational individuals, the payoff $\hat{x}_{-i}$ that the individual believes the other person will get is the true expected payoff $E(x_{-i})$. However, if the individual

is not rational, or does not consider expectations as an appropriate concept in one-shot situations, instead of $E(x_{-i})$ she can "hope for" any other value of $x_{-i}$ from $X$. This will be discussed in more detail below.

If the individual decides to learn about the consequences of her choices (index $U$ for "uncover") and anticipates the choice she will make once the corresponding payoffs to individual $-i$ are uncovered, she expects utility

$$EU_U(x_i, x_{-i}) = E[u(x_i) - D(f_U^{-1}(x_i, x_{-i}))] \quad . \tag{4}$$

The function $f_U$ now maps standards to choices given the choice set $X_U$ that assigns an $x_{-i}$ to each $x_i$ under voluntary disclosure. If voluntary and enforced disclosure are equal in the perception of the individual, $f_U = f_F$ and $EU_U(x_i, x_{-i}) = EU_F(x_i, x_{-i})$, given that the individual correctly predicts her choice. Otherwise the expected utilities may differ.

The difference between ignorance and disclosure is obvious: for the utility in (3), the individual knows what payoff she gets, and experiences dissonance derived from the payoff she believes the other person will get. For the utility in (4), she expects herself to make certain decisions given the available alternatives, which then induce utility and (possibly) dissonance. Knowing the probabilities of the different choice sets the individual can face, she can derive the expected utility and dissonance she will experience if disclosing the payoffs. Hence, she chooses to ignore the exact consequences of her decision if

$$u(x_i) - E[u(x_i)] > D(f_I^{-1}(x_i, \hat{x}_{-i})) - E[D(f_U^{-1}(x_i, x_{-i}))] \quad . \tag{5}$$

i.e., if the expected loss in utility from outcomes exceeds the difference between the dissonance experienced when anticipating the other person to receive the expected payoff, and the expected dissonance when making an informed decision.

If an individual - fully informed about the payoffs - expects herself to make decisions in compliance with her standards, $f_U^{-1}(E[x_i, x_{-i}]) \succeq \Delta^*$, condition (5) reduces to

$$u(x_i) - E[u(x_i)] > D(f_I^{-1}(x_i, \hat{x}_{-i})) \quad .$$

She chooses to ignore the consequences of her behavior, if the dissonance that results from this is lower than the loss in utility from outcomes that

she expects from behaving in accordance with her standards when making an informed decision.

The first distinction we make in the experiment is whether subjects are "selfish" or "social" in situations where the consequences of their choices are known. To check for the robustness of our results, categorizations are made on the basis of two different criteria: According to the first criterium, when faced with the choice of keeping 7 Euro for themselves and giving 1 Euro to the other, or keeping 4 Euro and giving 4 Euro to the other, selfish subjects choose (7,1) while social subjects choose (4,4). This means that

$$\begin{aligned} u(7) - D(f_{F1}^{-1}(7,1)) &> u(4) - D(f_{F1}^{-1}(4,4)) \quad \text{for selfish individuals and} \\ u(7) - D(f_{F1}^{-1}(7,1)) &< u(4) - D(f_{F1}^{-1}(4,4)) \quad \text{for social individuals,} \end{aligned}$$

with $X_{F1} = [(7,1),(4,4)]$. According to the second criterium, in the strategy method selfish subjects choose 10 Euro for themselves and 0 Euro for the other, rather than 6,50 Euro for themselves and 6.50 Euro for the other. In contrast, social subjects prefer 6.50 Euro for themselves and 6.50 Euro for the other over 10 Euro for themselves and 0 Euro for the other:

$$\begin{aligned} u(10) - D(f_{F2}^{-1}(10,0)) &> u(6.5) - D(f_{F2}^{-1}(6.5,6.5)) \quad \text{for selfish individuals and} \\ u(10) - D(f_{F2}^{-1}(10,0)) &< u(6.5) - D(f_{F2}^{-1}(6.5,6.5)) \quad \text{for social individuals,} \end{aligned}$$

with $X_{F2} = [(10,0),(6.5,6.5)]$.

For reasons of expositional clarity, in what follows we will refer only to the first criterium. However, the hypotheses we arrive at do not depend on the exact criterium with which we distinguish selfish from social individuals. In the analysis of the data in section 4, results on the basis of both criteria are presented.

Under the assumption that people do not experience dissonance when sharing equally with the other person, social people experience stronger dissonance when not sharing than selfish people, relative to their utility from monetary outcomes:

$$\begin{aligned} u(7) - u(4) &> D(f_{F1}^{-1}(4,4)) \quad \text{for selfish individuals while} \\ u(7) - u(4) &< D(f_{F1}^{-1}(4,4)) \quad \text{for social individuals.} \end{aligned}$$

Individuals that are categorized as selfish on the basis of these criteria may or may not ignore the consequences of their choice when payoffs $x_{-i}$ are not

visible initially. For the analysis of the reasons of other-regarding behavior, however, their choices are not relevant, simply because they do not show such behavior. Instead, we consider in detail the choices of those subjects who are categorized as "social" on the basis of the above criteria.

When the payoffs of individual $-i$ are not visible initially, social individuals prefer to ignore the consequences of their decision and take a payoff of 7 Euro for themselves rather than to uncover the consequences of their choice if

$$u(7) - D(f_I^{-1}(7, \hat{x}_{-i})) \quad > \quad E[u(x_i) - D(f_U^{-1}(x_i, x_{-i}))] \tag{6}$$

For the right hand term, the probabilities for the occurrence of the different relevant choice sets are

$$\begin{aligned}
X_{U1} &= [(7,1),(4,4),..] \quad \text{with} \quad p_{U1} = 1/3 \\
X_{U2} &= [(7,1),(4,1),(4,7),..] \quad \text{with} \quad p_{U2} = 1/15 \\
X_{U3} &= [(7,7),..] \quad \text{with} \quad p_{U3} = 1/5 \\
X_{U4} &= [(7,4),(4,4),(4,1),..] \quad \text{with} \quad p_{U4} = 2/15 \\
X_{U5} &= [(7,4),(4,7),(4,1),..] \quad \text{with} \quad p_{U5} = 2/15 \\
X_{U6} &= [(7,4),(4,1),(4,1),..] \quad \text{with} \quad p_{U6} = 1/15 \\
X_{U7} &= [(7,4),(4,7),(4,4),..] \quad \text{with} \quad p_{U7} = 1/15
\end{aligned}$$

Condition (5) can then be written as

$$u(7) - D(f_I^{-1}(7, \hat{x}_{-i})) \quad > \quad \sum_{U=U1}^{U7} p_U[u(x_{i,U}) - D(f_U^{-1}(x_{i,U}, x_{-i,U}))]$$

Hence, the individual ignores the consequences of her choices if

$$u(7) - \sum_{U=U1}^{U7} p_U[u(x_{i,U})] > D(f_I^{-1}(7, \hat{x}_{-i})) - \sum_{U=U1}^{U7} p_U[D(f_U^{-1}(x_{i,U}, x_{-i,U}))]$$

i.e., if the expected loss in monetary utility is higher than the expected decrease in cognitive dissonance.

Consider a social individual who expects herself to make the following choices if consequences are uncovered, which are the majority choices of social individuals observed in the experiment: $X_{U1} \Rightarrow (4,4)$, $X_{U2} \Rightarrow (7,1)$, $X_{U3} \Rightarrow (7,7)$, $X_{U4} \Rightarrow (7,4)$, $X_{U5} \Rightarrow (7,4)$, $X_{U6} \Rightarrow (7,4)$, $X_{U7} \Rightarrow (7,4)$, and not to experience cognitive dissonance from these choices. Such an individual chooses to ignore the consequences of her choice if

$$u(7) - \frac{1}{3}u(4) - \frac{2}{3}u(7) > D(f_I^{-1}(7, \hat{x}_{-i})$$

If she behaved socially in the full information case, $u(7) - D(f_{F1}^{-1}(7,1)) < u(4)$. This means that for social subjects who ignore the consequences of their behavior if this is possible,

$$3D(f_I^{-1}(7,\hat{x}_{-i})) < D(f_{F1}^{-1}(7,1)) \quad , \tag{7}$$

i.e., the dissonance they experience when ignoring the consequences of their choice is less than a third of the dissonance when choosing (7,1) over (4,4) in the full information situation. In contrast, for social individuals who uncover the information on the outcomes of the other individual

$$3D(f_I^{-1}(7,\hat{x}_{-i})) > D(f_F^{-1}(7,1)) \quad . \tag{8}$$

Ignorance by social individuals can therefore have two complementary reasons: First, $f_I$ can be "flatter" than $f_F$, that is, if ignorance is possible, the same behavioral standard permits a choice which focuses more on the individual's own outcome. Second, $\hat{x}_{-i}$ can be very high compared to 1. Setting $\hat{x}_{-i} = E(x_{-i})$ under rational expectations may be sufficient for (8) to be fulfilled. However, the higher an individual believes $\hat{x}_{-i}$ to be, e.g., by focusing on the outcomes 4 and/or 7, the lower her cognitive dissonance, and the more likely it is that she will ignore the consequences of her choice.

If a significant number of people in the experiment show other-regarding behavior if the consequences of their choices are uncovered, but ignore these consequences if given the chance, this means that part of the observed "fair" behavior is not driven by genuine preferences for equality, but is consistent with the avoidance of cognitive dissonance. Even the people who show other-regarding behavior and do not ignore the social dilemma situation do not necessarily have genuine preferences for equality. They may as well have preferences for a higher standard of behavior, or a mapping function which results in higher dissonance given a certain choice. Similarly, people that never show other-regarding behavior, instead of having a genuine preference for selfishness, may have a preference for a different standard, e.g., being smart, or a mapping function which results in lower dissonance given a certain choice.

### 3.4 Hypotheses

From the above we can derive the following hypotheses for the experiment.

**Hypothesis 1** *Social individuals who ignore the consequences of their behavior have more difficulty on average to decide whether to uncover the consequences of their choice in periods 1 and 3.*

Selfish individuals do not find it difficult to make this decision, since they know that in the end they will choose 7 Euro for themselves in any case. They face a tradeoff between monetary payoff and cognitive dissonance, which for the average selfish individual is clearly in favor of the monetary payoff, making the decision easy. Genuine social individuals would experience high cognitive dissonance if they ignored the consequences of their behavior. Hence, they face a tradeoff between monetary payoff and cognitive dissonance, which for the average genuine social individual is clearly against the monetary payoff, making the decision easy, too. For social ignorers, in contrast, the trade-off between monetary utility and dissonance is not that clear, because their dissonance functions lay in between the two extremes: they would experience more dissonance than the selfish types if they ignored, but less than the genuine pro-socials. Accordingly, for the average social ignorer it is harder to make this decision than for the other two types.

**Hypothesis 2** *The lower subjects believe the share of others is, who care about the other's payoff and do not ignore, the lower the probability that they, too, consider the other's payoff.*

This hypothesis results from general findings on the impact of social comparison, which suggest that a behavior seems more acceptable if it is conducted by a higher number of others in the reference group.

**Hypothesis 3** *The higher pro-social subjects "believe" $\hat{x}_{-i}$ to be, the higher is the probability of ignorance.*

If a subject who in general has a preference for behaving fairly can make herself believe that the individual affected by her choice receives a better payoff, she derives less dissonance from her behavior. Accordingly, the probability that she prefers ignorance over disclosure increases.

**Hypothesis 4** *If social subjects believe that ignorance is not unfair, the probability of ignorance increases.*

This hypothesis is derived directly from the mapping function, and the assumption that social individuals perceive fairness as a desirable behavioral standard.

# 4 Results

Our first interest is whether we can in fact observe "ignoring" behavior among individuals. In a first step we use data from period 4 of the experiment in order to classify the 99 participants into selfish and pro-social types. This classification is based on the strategy method-induced preferences for the decision between a (10,0) split and an equal split of (x,x), where x increased in steps of €0.5. €6.5 is used as the cutoff level to distinguish between selfish (always choice of (10,0) if $x \leq 6.5$) and pro-social (always choice of the equal split if $x > 6.5$) participants. This categorization produces 39 selfish and 60 pro-social types. Alternatively, data from the binary choices (for instance (7,1) vs. (4,4)) in period 4 can be used to classify participants. However, this data is not available for all participants since these questions were not used in all sessions. Hence, we focus our analysis on the classification based on the strategy method. The following results are robust with respect to the chosen categorization.
Next we are interested in the behavior of these pro-social types, when they face a situation as in period 1 and 3 (consequences of choices for the other participant are unknown, yet a mouse click will reveal them). Genuine pro-socials will always uncover the additional information so they can consider the other participant's payoff in their decision. If a pro-social participant does not uncover in either period 1 or 3, then we classify him/her as an "ignoring pro-social". We find 46 ignoring pro-socials and 14 genuine pro-socials.

With the participants classified into the types pro-selfs, genuine pro-socials and ignoring pro-socials, we run a multinomial probit regression. The base outcome is genuine prosocial, robust standard errors are clustered by sessions as the independent units of observation.

We used data from the experiment as well as answers in the post-experimental questionnaire as independent variables in the Probit regression. The questionnaire contained a standard set of emotion variables where participants

Table 1: Multinomial Probit with Type as Dependant Variable

| | ignoring pro-socials | | pro-selfs | |
|---|---|---|---|---|
| | coeff. | st.error | coeff. | st.error |
| Conflict | 1.276 | .3331 *** | .4636 | .4895 |
| Time | .0416 | .0144 *** | .0311 | .016 * |
| Change | .4404 | .3475 | -.3536 | .3842 |
| ShareIgnorers | 3.567 | 1.183 *** | 5.113 | 1.122 *** |
| angry | -.7241 | .4451 | -.292 | .326 |
| guilty | -1.1561 | .52 ** | -1.103 | .5656 * |
| proud | -1.25 | .4679 *** | -.7983 | .5579 |
| ashamed | 1.622 | .7289 ** | 1.703 | .8227 ** |
| OtherOpinion | .4836 | .4198 | .7018 | .3903 * |
| Age | -.0394 | .0981 | .1153 | .1688 |
| female | .3803 | .9282 | .6828 | .7561 |
| constant | -3.532 | 3.760 | -7.388 | 5.103 |

Log pseudolikelihood = -66.437762

indicated how they felt on a scale from 1 to 5. We also asked for background information like age, gender, their tendency to buy biological/organic or fair trade products, how relevant the opinion of others is to them and what they believe is the share of ignorers among the other participants. The variable "Change" means a participant switched from one option in period 1 to another in period 2. Remember that payoff couples were equal in these periods. The difference between the rounds is that payoffs of the other participant were hidden in period 1 (but it was possible to uncover), while they were known in period 2. Finally, we computed the reaction times of participants.[6] These are the times a participant spent in front of the first screen (possible decisions: uncover payoffs of participant Y, uncover the decision table of participant X or choose an option (A to F) straightaway) and in front of the second screen (possible decisions: choice of A to F with the payoffs of Y uncovered or the decision table of X shown).

**Hypothesis 1** The Probit regression shows positive and significant coeffi-

[6]Rubinstein (2007) propagates the use of response times in order to explore the deliberation process of decision makers. A positive correlation between the experienced conflict and response time has been found by Branas-Garza et al. (2007) among others.

cients for the conflict variable[7] and the reaction time (both at the 1%-level) for ignoring pro-socials.

It appears that participants classified as ignoring pro-socials experience a conflict. They also tend to require more time when they face the period 1/3 decisions (whether to uncover and enter a social dilemma or select one of the options right away). These results are in line with the prediction of the cognitive dissonance-based model. Participants who otherwise act socially-minded do not uncover (our ignoring pro-socials type). They should experience a high level of cognitive dissonance due to the two conflicting psychological cognitions: A general sense for sharing equally that was however dismissed in a situation where ignoring the consequences for the other was facilitated and a material advantage was gained at the expense of the other. Such a dissonance between these two inconsistent cognitions should result in experiencing a conflict and in longer times to decide between the options. Our experimental results confirm this.

**Hypothesis 2** Participants' beliefs about the share of others who ignore seems to explain selfish behavior given the 1% significance level in the regression for the pro-self type. Likewise, the effect for ignoring pro-socials is positive and significant at 1%.

**Hypothesis 3** Data about what participants believe the other participant will receive, $\hat{x}_{-i}$, is not available for all observations. The respective question has not been included in all sessions, participants who uncovered could not be used and some participants did not provide precise answers. In total, 38 observations were available, 22 of them were from "ignoring pro-socials". Their average belief $\hat{x}_{-i}$ was 3.46, greater than the expected value of 3.4, but not statistically significant. The pro-selfs' average was 3.27.

**Hypothesis 4**: Data about participants' own standard in situations 1 and 3 was collected in some of the sessions. Answers were not enforced, though. This textual data is not quite in line with observed behavior as 21 "ignoring pro-socials" wrote that they find ignorance acceptable, while 10 wrote one should actually uncover. This may have two reasons. First, answers to questionnaires were given with hindsight. Second, holding a perception and openly (though anonymously) stating it are two different things. To-

[7]The precise (translated) text of this question is: "Did you find it easy in these situations [period 1 or 3], whether you want to receive information about the payoff of participant Y or the payoff table of participant X?"

gether, these reasons may have led some subjects to state standards which are different from the ones they based their decisions on.

# 5 Conclusion

There exists extensive laboratory and field evidence on other-regarding behavior. This behavior has usually been explained with people having genuine other-regarding preferences. Recent experimental evidence on sorting and ignorance in social dilemma situations, including the evidence presented in this paper, has cast some doubt on this explanation. However, no theory which explains this evidence has been put forward so far.

We develop a new model, which can explain our data and that of earlier experiments. The model is based on cognitive dissonance, i.e., on the assumption that people do not show other-regarding behavior because they truly care for others, but because they want to maintain a certain self-perception. Using a within-subject design, we test the main predictions of our model and find support for them in the data. Note that also Houser et al. (2008) refer to - but not model - cognitive dissonance to explain their experimental data.

In spite of this supporting evidence, it must be emphasized that the experiment cannot serve as a full-fledged test of the model. More research is needed to establish the results and fine-tune its predictions. For the moment we suggest the model as an alternative to the known (outcome-based and intentions-based) social preferences models and also to more recent social esteem models following Benabou and Tirole (2006). It remains for further research to compare our suggested cognitive dissonance-based model to the existing approaches in order to find out more about the true determinants of human behavior.

## References


Aronson, E. (1997). Review: Back to the Future: Retrospective Review of Leon Festinger's 'A Theory of Cognitive Dissonance' *American Journal of Psychology*, **110**, No. 1 (Spring 1997), pp 127-137

Benabou, R. and J. Tirole (2006). Incentives and Pro-Social Behavior *American Economic Review*, **96(5)**: 1652-1678

Bolton, G. and A. Ockenfels (2000). A theory of equity, reciprocity and competition *American Economic Review*, **100**, 166-193

Branas-Garza, P., A. Leon-Mejia and L. M. Miller (2007). Response Time under Monetary Incentives: the Ultimatum Game *Jena Economic Research Papers, working paper 2007-070*.

Broberg, T., Ellingsen, T., and Johannesson, M. (2007). Is generosity involuntary? *Economics Letters*, **94**, 32–37.

Camerer, C. (2003). Behavioral Game Theory: Experiments in Strategic Interaction *Princeton University Press*, Princeton

Dana, J., Cain, D.M., and Dawes, R.M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games *Organizational Behavior and Human Decision Processes*. 100 (2006) 193-201.

Dana, J., Weber, R., and Kuang, J.X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness *Economic Theory*. 33 (2007) 67-80.

Dufwenberg, M., and G. Kirchsteiger (2004). A Theory of Sequential Reciprocity *Games and Economic Behavior*. 47, 268-298.

Falk, A., and U. Fischbacher (2005) A Theory of Reciprocity *Games and Economic Behavior*

Fehr, E. and K. Schmidt (1999). A Theory of Fairness, Competition and Cooperation *Quarterly Journal of Economics*, **114**, 817-868

Fehr, E. and K. Schmidt (2003) The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories *Handbook of Reciprocity, Gift-Giving and Altruism*

Festinger, L. (1957) A Theory of Cognitive Dissonance *Row Petersen*, Evanston, IL

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

Greiner, B. (2004). An Online Recruitment System for Economic Experiments. In K. Kremer and V. Macho, editors, *Forschung und wissenschaftliches Rechnen 2003*. GWDG Bericht 63, Gttingen : Ges. fr Wiss. Datenverarbeitung.

Houser, D., E. Xiao, K. McCabe and V. Smith (2008). When punishment fails: Research on sanctions, intentions and non-cooperation *Games and Economic Behavior*, **62**, 509–532.

Lazear, E., U. Malmendier and R. Weber (2006). Sorting in Experiments with Application to Social Preferences *NBER working paper*. No. W12041 (2006)

Matthey, A. and T. Regner (2007). Is observed other-regarding behavior always genuine? *Jena Economic Research Papers, working paper 2007-109*.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, **83**(5), 1281–1302.

Rabin, M. (1994). Cognitive dissonance and social changes. *Journal of Economic Behavior and Organization*, **23**, 177–194.

Rubinstein, A. (2007). Instinctive and cognitive resasoning: A study of response times. *Economic Journal*, **117** (523), 1243–1259.

Steele, C. M. (1988) The psychology of self-affirmation: Sustaining the integrity of the self *in: L. Berkowitz (ed.), Advances in experimental social psychology*, **21** pp 261-302), Academic Press, New York

# A Experimental Instructions

Welcome and thanks for participating in this experiment.

In this experiment you can win a certain amount of money, which depends on your and the other participants' decisions in the experiment. It is, therefore, important that you read the following instructions carefully.

Please note that these instructions are only meant for you and that you are not allowed to exchange any information with the other participants. Similarly, during the entire experiment it is not allowed to talk to the other participants. If you have any questions or concerns, please raise your hand. We will answer your questions individually. Please do not ask your questions aloud. It is very important that you follow these rules, since otherwise we have to stop the experiment. Please also turn of your mobile phones now.

## Procedure

The experiment lasts about 30 min. It consists of four parts. In each part you make decisions. Each decision will be explained again briefly on the monitor. While you make decisions, the other participants also make decisions which may influence your payoffs.

Your payoff from the experiment depends on your decision and the decisions of the other participants. But only one of the four parts is chosen randomly and your are paid in cash according to the payoff from this part. The exact procedure according to which your payoff is calculated is explained below. All amounts in the decision situation are given in Euro. They are paid out in exactly the given amount at the end of the experiment. In addition you receive 2.50 Euro as a show-up fee.

After you filled in a questionnaire the experiment ends and you receive your payoff.

Again the procedure as an overview:

- Reading of the instructions, test questions
- Four parts with decision situations
- Questionnaire
- Payoffs and end of the experiment

## Decision situations

This experiment consists of four parts in which you make decisions. In each of these parts you are randomly matched with two other participants of the experiment. This means that your decision influences the payoffs of one other participant. In turn, the decisions of yet another participant influence your decisions. You can be sure not to meet the same participants twice.

### Parts 1-3:

In the decision situations of parts 1-3 you can choose between five options. Each option includes a payoff for you and for one other participant. This participant is not known to you. In the experiment, he or she is simply called participant "Y". You can decide which option to choose, and with this choice decide upon the payoff four yourself and participant Y.

In all decision situation, the payoffs 1, 1, 4, 4, and 7 are assigned to the options A, B, C, D and E. Only the numbers 1, 1, 4, 4, and 7 can occur as payoffs, both for yourself and for participant Y. How numbers/amounts are assigned to options is determined randomly!

For example, assume that your and participant Y's payoffs are assigned to options according to the following table (in EUR):

| Option | A | B | C | D | E |
|---|---|---|---|---|---|
| You | 1 | 4 | 4 | 7 | 1 |
| Participant Y | 4 | 1 | 4 | 1 | 7 |

The choice of the options (A, B, C, D or E) that determines the payoffs for you and participant Y is up to you. Note that the above assignment of payoffs to options is only an example. It does not necessarily occur in the experiment.

It is important to understand that each participant receives his or her own randomly determined payoff table. With your choice of an option you determine the payoff for you and participant Y. In addition, another participant, who is called "X" in the experiment, determines the payoff for herself/himself and you through her/his own payoff table.

Which one of the two payoffs of each participant is paid out in the end - the one from his/her own decision or the one from the decision of his/her participant X, is determined randomly. The following graphic illustrates again the choice of the payoffs. Note that the participants X and Y you are matched with are different in each part of the experiment. This means you will not "meet" a given participant twice.

[Here followed a graphic that illustrated the matching of participants]

**Hidden payoffs**

In some decision situations of parts 1-3 of the experiment you are informed about your own payoffs from the options A, B, C, D and E, but not about the payoffs of participant Y. On the screen you will see the following table:

| Option | A | B | C | D | E |
|---|---|---|---|---|---|
| You | 1 | 4 | 4 | 7 | 1 |
| Participant Y | ? | ? | ? | ? | ? |

In this case you only know that the payoffs of participant Y are distributed uniformly. It is important that you understand what is meant by this.

As said above, in all decision situations the payoffs 1, 1, 4, 4, and 7 are assigned to the options A, B, C, D and E. Only the numbers 1, 1, 4, 4, and 7 can occur as payoffs. In each distribution 1, 1, 4, 4, and 7 occur for sure. This means twice the payoff 1, twice the payoff 4 and once the payoff 7. The combinations of options and payoffs are uniformly distributed, which means all occur with equal probability. As a consequence, for each option of your payoff table the payoff for participant Y amounts to 1 with probability 40%, to 4 with probability 40% and to 7 with probability 20%.

For example, two of the 30 possible payoff combinations for participant Y are:

| Option | A | B | C | D | E |
|---|---|---|---|---|---|
| You | 1 | 4 | 4 | 7 | 1 |
| Participant Y | 4 | 7 | 4 | 1 | 1 |

and

| Option | A | B | C | D | E |
|---|---|---|---|---|---|
| You | 1 | 4 | 4 | 7 | 1 |
| Participant Y | 1 | 4 | 7 | 1 | 4 |

**Possibility to uncover the payoffs**

If the payoffs of participant Y are hidden, e.g., you see the following table

| Option | A | B | C | D | E |
|---|---|---|---|---|---|
| You | 1 | 1 | 4 | 4 | 7 |
| Participant Y | ? | ? | ? | ? | ? |

You have several possibilities:

- You can directly choose one of the options A, B, C, D or E.
- You can uncover the payoffs of participant Y. You can do this simply by clicking on the button "Participant Y" on the screen. You do not have to uncover, but can also choose an option directly. If you decide to uncover and click on "Participant Y", the payoffs of participant Y are shown. The assignment of your payoffs and the payoffs of participant Y does not change when you uncover. Neither does it cost anything. Afterwards you can choose one of the options A, B, C, D or E. By uncovering you neither receive new information about your own payoffs, nor about the payoff table of player X, the one who decides about your payoff.

- You can uncover the payoff table of participant X. You can do this simply by clicking on the button "Participant X" on the screen. You do not have to uncover, but can also choose an option directly. If you decide to uncover the payoff table of participant X, you see the payoffs of participant X and your own payoffs that participant X decides upon according to his payoff table. You cannot see which option participant X chooses! Uncovering the payoff table of participant X does not cost you anything either. You cannot receive any information regarding the payoffs of participant Y once you have chosen to uncover the payoff table of participant X. Afterwards you can just decide which of the options A, B, C, D or E to choose.

**Part 4**

In this part you also make decisions that are relevant for your own payoffs and the payoffs of participant Y. (Reminder: Participants X and Y are different in each part of the experiment.) For each decision you have to choose between two pairs of payoffs. Each pair of payoffs assigns one payoff to you and one to participant Y. You have to decide which of the payoff pairs you prefer. The particular pairs of payoffs you have to choose between are shown to you on the screen.

**Overview over the most important features of the experiment:**

- You make decisions in four parts of the experiment. In each part you are matched with two different participants X and Y.
- Your decisions also apply to the player who is participant Y in this part of the experiment.
- The player who is participant X in this part of the experiment makes his decisions (based on his own payoff table!), which assigns a payoff also to you.
- For the decision situations in parts 1-3 of the experiment you choose between five options. Your payoffs are uniformly distributed across options and always known to you. The payoffs of participant Y are uniformly distributed and not always known, but they can always be

uncovered. You can choose one of the five options and with this choice determine the payoffs for yourself and participant Y.

- In part 4 of the experiment you choose between pairs of payoffs, that each assign one payoff to yourself and one to participant Y.

### Your cash payoff from the experiment

Your payoff from the experiment depends on your decision and the decisions of the other participants that you are matched with randomly in the four parts of the experiment. But only one of the four parts is chosen randomly and your are paid in cash according to the payoff from this part.

To determine this part and the decision that is relevant for your payment (yours or that of participant X), you have to make two more choices after part 4 of the experiment. For the first choice you choose between 8 letters. Each letter stands for one part (four parts) and one of the two "roles". With role we mean whether your decision or that of participant X is relevant for your payment. The order of the letters has no meaning related to the parts and roles. If you choose a letter which assigns to you a payoff from part four of the experiment, you have to make another decision (only then!). The reason is that in this part you make several decisions. For this second decision you again choose between letters, that is, you choose which of the decisions of part 4 is relevant for your payment. Again, the order of the letters has no meaning related to the decision that is relevant for you. In order for you to be able to check whether your choice is consistent with your payment, we provide an envelop with the assignment of letters to parts and roles and depose it well in sight. You can open it after the experiment.

With this procedure all parts of the experiment have the same chance of being relevant for your payment. In addition, the chance that your own decision is relevant for your payment is just as big as the chance that the decision of another participant is relevant for your payment. The same is true for all other participants.

Your payment is prepared while you complete the questionnaire. You receive your payment and sign a receipt before you leave the laboratory.

If you have questions regarding the content or procedure of the experiment, please raise your hand now. Then answer the following questions to ensure

that you understood the instructions. Only if you answer these questions correctly you can participate in the experiment.