

Matthey, Astrid; Regner, Tobias

**Working Paper**

## Is observed other-regarding behavior always genuine?

Jena Economic Research Papers, No. 2007,109

**Provided in Cooperation with:**

Max Planck Institute of Economics

*Suggested Citation:* Matthey, Astrid; Regner, Tobias (2007) : Is observed other-regarding behavior always genuine?, Jena Economic Research Papers, No. 2007,109, Friedrich Schiller University Jena and Max Planck Institute of Economics, Jena

This Version is available at:

<https://hdl.handle.net/10419/25684>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



# JENA ECONOMIC RESEARCH PAPERS



# 2007 – 109

## **Is observed other-regarding behavior always genuine?**

by

**Astrid Matthey  
Tobias Regner**

[www.jenecon.de](http://www.jenecon.de)

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact [m.pasche@wiwi.uni-jena.de](mailto:m.pasche@wiwi.uni-jena.de).

### Impressum:

Friedrich Schiller University Jena  
Carl-Zeiss-Str. 3  
D-07743 Jena  
[www.uni-jena.de](http://www.uni-jena.de)

Max Planck Institute of Economics  
Kahlaische Str. 10  
D-07745 Jena  
[www.econ.mpg.de](http://www.econ.mpg.de)

© by the author.

# Is observed other-regarding behavior always genuine?

Astrid Matthey ♠ Tobias Regner ♣

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

♣ *Friedrich Schiller University, Jena, Germany*

January 17, 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), people can choose whether to learn about the consequences of their choice for the receiver. We find that a majority of the people that show other-regarding behavior when the payoffs of the receiver are known chose to ignore them if possible. This behavior is inconsistent with genuine other-regarding 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:* C9, C7, D8

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

# 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.<sup>1</sup> 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".

While our participants make choices in standard social dilemma situations (that is, games with an open payoff set), we also introduce the possibility to remain uninformed about the other participant's payoff. Our within-subjects experimental design contains situations where people know that a social dilemma exists, i.e., that their behavior affects others. But they can also 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.

The effect of varying levels of information in social dilemma situations has been studied before. Our design differs from the works of, e.g., Dana et al (2005), Lazear et al (2006) and Broberg et al (2007), where people can avoid entering the social dilemma situation at all. Our main interest is rather whether there is a change of participants' behavior given that they stay in a social dilemma: Do people who show other-regarding behavior given the open payoff set choose to find out the consequences of their action if they can remain uninformed? In other words, are they genuinely pro-social? Or do they act pro-social with an open payoff set and prefer to remain ignorant when given the chance (taking negative consequences for the other participants into account)? Furthermore, we are also interested in the driving motivational forces of such "ignorers" that distinguish them from genuinely pro-socials.

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

Participants in our experimental setup may be in such a situation. When participants prefer a fair split under full information but choose to ignore under limited information, they will have two inconsistent psychological cognitions. Hence, they should experience cognitive dissonance. Our cognitive dissonance-based model explains the observed split into genuine and ignoring pro-socials, while other social preferences models cannot. Finally, our experimental design enables us to test the suggested model. The study of Dana et al. (2007) supports our experimental results, though they use a setup that does not allow the test of a theory. Ours does.

---

<sup>1</sup>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))

A motivating example for the analysis of such situations is the fact that they reflect the features of many decisions we make every day. Whenever we shop for food, clothes etc., we face the decision of whether to pay higher prices for fair trade and/or organic products, i.e. products of firms which maintain high social and/or ecological standards. This means that we decide whether to share our money with others, by paying more than the cheapest available price of the good. 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 would like to simulate in our experiment. In contrast, the sorting experiments of Dana et al. (2005) and Lazear et al. (2006) offer people an alternative where their behavior does not have an 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, cloths or other consumption goods. Neither is it available in many social interactions which do not involve money, e.g., when driving home during rush hour. There we have to decide whether we want to save only on our own time by driving aggressively, or to share the time saved with others by driving courteously. If we prefer to avoid this decision we can ignore the effect of our driving on others, and just get home as fast as we can. However, we cannot avoid the social dilemma situation itself.

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 could 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) chose an option immediately, i.e., without knowing the consequences for Y, ii) disclose the outcomes that each option assigned to participant Y, and then chose an option, or iii) see the outcome pairs from which participant X could chose, and then chose 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 disclosed the outcomes of participant Y in periods one and three, how much they thought "their" participant Y received if they did not disclose his outcomes, and what they think 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. 63 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 €7.54. The show-up fee for the experiment amounted to €2.5.

At their 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.

### 3.1 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

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

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, 2005) 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.<sup>2</sup>

In this paper we chose 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, 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.

---

<sup>2</sup>Due, presumably, to the different focus of the authors.

In this sense, people may have a preference regarding "behavior towards 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 in addition 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 which deviates from the standard by which they measure themselves, they experience cognitive dissonance (e.g., Aronson, 1997). Since this results in negative emotions, they like to avoid such deviations. This can explain standard other-regarding behavior, but also the behavior we observe in the experiment, which is not consistent with earlier models. Our experimental results are supported by Dana et al (2007), who find similar effects.

### 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  $(x_i, x_{-i1}, \dots, x_{-in})$ , 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  $j = 1, \dots, m$  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  $\mathcal{X}$ .

Next, define  $E_{i,-i}(X)$  as the expectation individual  $i$  thinks individual  $-i$  has regarding the choice of individual  $i$ , given the choice set  $X$ :  $E_{i,-i}(X) := E_i[E_{-i}(x_i, x_{-i})]$ . In the more general case,  $E_{i,-i}(X)$  would not only include the expectation of individual  $-i$ , but the expectations of all relevant others regarding the choice of individual  $i$ .  $E_{i,-i}(X)$  would then form some kind of social norm. However, since social norms are neither the focus of this paper, nor reflected in the experimental design, we restrict  $E_{i,-i}(X)$  to focus only on the expectations of the individual that is affected by  $i$ 's choice.

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  $\Delta$  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.<sup>3</sup>

We then define a function  $f$  which maps behavioral standards into choices, dependent on the available choice set, the individual's expectation of others' choices, and her perception of what the individual who is affected by her choice expects her to choose:  $f(\Delta, X, \mathcal{X}, E_{i,-i}(X)) : \Delta \times X \times X \times X \rightarrow X$ .

This means that if  $X$ ,  $\mathcal{X}$  and  $E_{i,-i}(X)$  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, 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. Finally,  $E_{i,-i}(X)$  is included in the mapping function to account for evidence which suggests that what people think others expect of them effects their behavior (Dana et al, 2006).

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

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

$u(x_i)$  is the utility that the individual derives from what she herself receives as the consequence of her decision. The general form of this utility function is similar to the model of Rabin (2003). In contrast to his model, however, we analyze cognitive dissonance in detail.  $D(\Delta^*, \Delta)$  is the cognitive dissonance she experiences from her choice. This dissonance depends on how the behavioral standard  $\Delta$  that is consistent with her behavior compares to her desired behavior  $\Delta^*$ .<sup>4</sup>

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

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:  $D(\Delta^*, f^{-1}(x_i > 50\%, x_{-i} < 50\%)) > 0$ .

---

<sup>3</sup>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.

<sup>4</sup>This ignores possible pride from behavior which is consistent with a standard higher than  $\Delta^*$ , i.e.,  $D < 0$ .

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 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 chose 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). The functions which apply to these situations receive the index  $K$  for "known" information. Choice  $(x_i, x_{-i})$  given choice set  $X_K$  then induces utility

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

The function  $f_K$  maps standards to choices, given the choice set  $X_K$  that assigns an  $x_{-i}$  to each  $x_i$  the individual may choose. Given this choice set, the individual forms  $\mathcal{X}_K$  and  $E(X_K)$ .

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). Her utility in case of ignorance (index  $I$ ), i.e., choosing an alternative without learning about the consequences, is then given as

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

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 disclosed 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  $X$  and voluntary ignorance. Again, given  $X_I$  the individual will form a set of expectations  $\mathcal{X}_I$  over the choices of others and a belief of what individual  $-i$  expects her to do:  $E_{i,-i}(X_I)$ .

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 disclosed, she expects utility

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

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,  $f^d = f^o$  and  $EU_U(x_i, x_{-i}) = EU_K(x_i, x_{-i})$ , 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

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

i.e., if the expected loss in utility 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 with open information about the payoffs an individual expects herself to make decisions in compliance with her standards,  $f_U^{-1}(E[x_i, x_{-i}]) \succeq \Delta^*$ , condition (6) 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 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

$$\begin{aligned} u(7) - D(f_{o1}^{-1}(7, 1)) &> u(4) - D(f_{o1}^{-1}(4, 4)) \quad \text{for selfish individuals and} \\ u(7) - D(f_{o1}^{-1}(7, 1)) &< u(4) - D(f_{o1}^{-1}(4, 4)) \quad \text{for social individuals,} \end{aligned}$$

with  $X_{o1} = [(7, 1), (4, 4)]$ . According to the second criterium, in the strategy method selfish subjects choose 10 Euro for themselves and 0 Euro for the other, rather than 5 Euro for themselves and 5 Euro for the other. In contrast, social subjects prefer 5 Euro

for themselves and 5 Euro for the other over 10 Euro for themselves and 0 Euro for the other:

$$\begin{aligned} u(10) - D(f_{o2}^{-1}(10, 0)) &> u(5) - D(f_{o2}^{-1}(5, 5)) && \text{for selfish individuals and} \\ u(10) - D(f_{o2}^{-1}(10, 0)) &< u(5) - D(f_{o2}^{-1}(5, 5)) && \text{for social individuals,} \end{aligned}$$

with  $X_{o2} = [(10, 0), (5, 5)]$ .

For reasons of expositional clarity, in what follows we will refer only to the first criterium. However, the hypotheses we arrive 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:

$$\begin{aligned} u(7) - u(4) &> D(f_{o1}^{-1}(4, 4)) && \text{for selfish individuals while} \\ u(7) - u(4) &< D(f_{o1}^{-1}(4, 4)) && \text{for social individuals.} \end{aligned}$$

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 disclose the consequences of their choice if

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

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

$$\begin{aligned} X_{d1} &= [(7, 1), (4, 4), ..] && \text{with } p_{d1} = 1/3 \\ X_{d2} &= [(7, 1), (4, 1), (4, 7), ..] && \text{with } p_{d2} = 1/15 \\ X_{d3} &= [(7, 7), ..] && \text{with } p_{d3} = 1/5 \\ X_{d4} &= [(7, 4), (4, 4), (4, 1), ..] && \text{with } p_{d4} = 2/15 \\ X_{d5} &= [(7, 4), (4, 7), (4, 1), ..] && \text{with } p_{d5} = 2/15 \\ X_{d6} &= [(7, 4), (4, 1), (4, 1), ..] && \text{with } p_{d6} = 1/15 \\ X_{d7} &= [(7, 4), (4, 7), (4, 4), ..] && \text{with } p_{d7} = 1/15 \end{aligned}$$

Condition (6) can then be written as

$$u(7) - D(f_I^{-1}(7, \hat{x}_{-i})) > \sum_{d=d1}^{d7} p_U [u(x_{i,d}) - D(f_U^{-1}(x_{i,d}, x_{-i,d}))]$$

Hence, the individual ignores the consequences of her choices if

$$u(7) - \sum_{d=d1}^{d7} p_U [u(x_{i,d})] > D(f_I^{-1}(7, \hat{x}_{-i})) - \sum_{d=d1}^{d7} p_U [D(f_U^{-1}(x_{i,d}, x_{-i,d}))]$$

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 disclosed, which are the majority choices of social individuals observed in the experiment:  $X_{d1} \Rightarrow (4, 4)$ ,  $X_{d2} \Rightarrow (7, 1)$ ,  $X_{d3} \Rightarrow (7, 7)$ ,  $X_{d4} \Rightarrow (7, 4)$ ,  $X_{d5} \Rightarrow (7, 4)$ ,  $X_{d6} \Rightarrow (7, 4)$ ,  $X_{d7} \Rightarrow (7, 4)$ , 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 open case,  $u(7) - D(f_{o1}^{-1}(7, 1)) < u(4)$ . This means that for social subjects who ignore the consequences of their behavior if this is possible,

$$3D(f_I^{-1}(7, \hat{x}_{-i})) < D(f_{o1}^{-1}(7, 1)) \quad ,$$

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 open situation. In contrast, for social individuals who disclose the information on the outcomes of the other individual

$$3D(f_I^{-1}(7, \hat{x}_{-i})) > D(f_{o1}^{-1}(7, 1)) \quad .$$

Ignorance by social individuals can therefore have two complementary reasons: First,  $f_I$  can be "flatter" than  $f_K$ , 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 (7) 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 disclosed, 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 disclose 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 selfishs 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 higher 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 received a better payoff, she derives less dissonance from her behavior. Accordingly, the probability that she prefers ignorance over disclosure increases.

**Hypothesis 3** *The lower subjects believe the share of others is, who care about the other's payoff and do not ignore, the higher 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 4** *If social subjects believe that ignorance is not unfair, 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 that otherwise elicit social preferences. In a first step we use data from period 4 of the experiment in order to classify the 63 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 categorisation results in 23 selfish and 40 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

classification.

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). Genuinely pro-socials will always uncover the additional information so they can consider the other participants payoff in their decision. If a pro-social participant did not uncover in either period 1 or 3, then we classify him/her as an "ignoring pro-social". We find 30 ignoring pro-socials and 10 genuinely pro-socials.

With the participants classified into the types pro-self, genuinely pro-social and ignoring pro-social, we run a Probit regression with the type as the dependent variable:

Table 1: Probit with type as the dependent variable

	ignoring pro-socials		pro-selfs		genuine pro-socials	
	coeff.	st.error	coeff.	st.error	coeff.	st.error
conflict	1.2156	.477 ***	-.6563	.3589 *	-.6412	.7007
angry	-.3052	.5269	-.0926	.5608	-.2588	.5991
guilt	-.3395	.4714	-.1583	.4849	1.5625	1.3887
proud	-.3265	.3804	.4553	.3291	-.4381	.5614
ashamed	.1278	.4995	.7261	.5004	-	-
Bio	1.1622	.5517 **	-.6870	.3898 *	.2604	.6651
FairTrade	.0586	.3953	.4124	.3678	-.2105	.7061
OtherOpinion	.2816	.3288	.1114	.2874	-1.167	.8303
ShareIgnorers	-.3993	1.111	3.0911	1.144 ***	-6.7408	4.4481
Age	-.2208	.1438	.3526	.1503	-.2183	.3233
female	-1.206	.6772 *	1.2068	.6059 **	.0718	.6486
Change	1.381	.6193 **	-1.595	.6406 **	1.473	1.3509
Time1Table	.0240	.0239	.0187	.0188	-.0842	.0518
Time1Choice	-.0648	.0313 **	.0020	.0217	.1366	.0777 *
Time3Table	.0672	.0336 **	-.0292	.0268	-.1396	.1013
Time3Choice	.0392	.0347	-.0344	.0249	-.0612	.0711
constant	-.722	3.472	-9.4654	4.0319	10.7261	9.3199
$R^2$	0.5943		0.4676		0.5056	

We used data from the experiment as well as answers in the post-experiment questionnaire as independent variables in the Probit regression. The questionnaire contained a standard set of emotion variables where participants 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 in period 1 and period 3. These reaction times are split into the time the 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 for "ignoring pro-socials" shows positive and significant coefficients for the conflict variable (at the 1%-level) and the time it took at the first decision screen (in period 3, at the 5%-level). In addition Wilcoxon rank-sum tests confirm that "ignoring pro-socials" tend to take more time to make the decision at the first screen (1% significance level in period 1 and 5%-level in period 3). It appears that participants classified as "ignoring pro-socials" experience a conflict. They also tend to require more time when they face the decision whether to uncover and enter a social dilemma or selecting 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** 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, 17 observations were available, 14 of them were from "ignoring pro-socials". Their average belief  $\hat{x}_{-i}$  was 3.59, greater than the expected value of 3.4, but not statistically significant due to the low number of observations.

**Hypothesis 3** Participants' belief 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. The effects for "ignoring pro-socials" and "genuine pro-socials" go in the right direction, but are not significant.

**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 only 7 "ignoring pro-socials" wrote that they find ignorance acceptable, while 7 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. Together, these reasons may have led some subjects to state standards which are different from the ones they based their decisions on.

Finally, the results for the variable "Change" are as well in line with our theory. It indicates a switch from one option chosen in period 1 to another option in period 2. As the distribution of payoffs is equal in these two periods, such behavior is inconsistent with pro-self or genuinely pro-social behavior, while it is possible among "ignoring pro-socials". The high level of significance in the regression confirms that.



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.

In spite of this supporting evidence, it must be emphasized that the experiment is only a first test of the model. More research is needed to establish the results and fine-tune its predictions. For example, an exogenous variation of the factors that are expected to influence the mapping from behavioral standards into decisions would shed light on the relevance of each factor, and on possible context dependence. Only then could the model be used to predict behavior in particular situations.

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 determinants of human behavior.

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

Jena Economic Research Papers 2007-109  
**A Experimental Instructions**

Instructions for the experiment are available upon request from the authors.