Selfish in the end?

An investigation of consistency and stability of individual behavior

Jeannette Brosig[*],

University of Duisburg-Essen, Germany

Thomas Riechmann, and Joachim Weimann[†‡]

University of Magdeburg, Germany

Abstract:

This paper puts general versions of prominent specifications of 'other-regarding' preferences to the experimental test. In a series of experiments based on various dictator and prisoner's dilemma games, we compare these concepts and the classical selfish approach. The experiments are special with regard to two aspects: First, we investigate the *consistency* of individual behavior within and across different classes of games. Second, we analyze the *stability* of individual behavior over time by running the same experiments on the same subjects at several points in time. Our results demonstrate that, in the first wave of experiments, all notions of other-regarding preferences explain a high share of individual decisions. Other-regarding preferences seem to wash out over time, however. In the final wave, it is the classical theory of selfish behavior that delivers the best explanation. Stable behavior over time is observed only for subjects who behave strictly selfishly.

Keywords: individual preferences, consistency, stability, experimental economics

JEL classification: C91, C90, C72, C73

---

[*] University of Duisburg-Essen, Universitaetsstrasse 12, 45117 Essen, Germany; phone: ++49-(0)201-183-2479, fax: ++49-(0)201-183-2879, email: jeannette.brosig@ibes.uni-due.de.

[†] University of Magdeburg, Postfach 4120, 39016 Magdeburg, Germany; phone: ++49-(0)391-6718547, fax: ++49-(0)391-6712971, email: thomas.riechmann@ww.uni-magdeburg.de, weimann@ww.uni-magdeburg.de.

[‡] The authors thank Bernd Irlenbusch, Hans-Theo Normann,, Simon Gächter, Axel Ockenfels, Garry Bolton, and Karim Sadrieh for providing insightful comments and suggestions.

1 Introduction

For a long time, economic science was built on a specification of individual preferences that implied rational and purely self-interested behavior. However, experimental research has produced a number of stylized facts that cast some doubt on the empirical validity of this specification. Subjects make voluntary contributions in public good games (for an early overview see Ledyard, 1995), they cooperate in the prisoner's dilemma (as already shown in Flood, 1952, 1958), and they make significant donations to others in dictator games (see, e.g., the early studies by Kahneman, Knetsch, and Thaler, 1986, Forsythe, Horowitz, Savin, and Sefton, 1994). None of these findings can be fully accounted for by the standard approach of rational and selfish behavior.

One way to tackle this problem is to deviate from the assumption of pure self-interest while maintaining the rational choice hypothesis. This path has been followed by a number of theorists who integrate some kind of other-regarding behavior into the individual preference model in order to organize the experimental data. For example, the theories developed by Bolton and Ockenfels (2000) and by Fehr and Schmidt (1999) are based on the supposition that people are not only interested in their own absolute payoff, but also in their own *relative* payoff. Charness and Rabin (2003) propose a theory of social preferences that assumes that subjects care about their own payoff, the others' payoff, and about efficiency. Andreoni and Miller (2002) and Andreoni, Castillo, and Petrie (2005) model a concern for altruism and efficiency by defining utility functions for keeping for oneself and giving to others.[1]

Fisman et al. (2007) use the modified dictator games introduced by Andreoni and Miller (2002) to check whether behavior that is not purely selfish can be rationalized by the assumption that subjects optimize their choice between payoff to *self* and payoff to *other*. They find that, first, individual behavior can be rationalized to a great extent with a well-behaved utility function and that, second, there is great heterogeneity among subjects regarding the shape of this function.

Beside the attempts to describe behavior in dictator games by assuming rational and non-selfish behavior, List (2007) offers an alternative approach. In a series of experiments, List alters the

[1] Alternative approaches assume some kind of reciprocity caused by the harming or helping intentions of the fellow players (see, e.g., Geanakoplos, Pearse, and Stacchetti, 1989, Rabin, 1993, Levine, 1998, Dufwenberg and Kirchsteiger, 1998, Falk and Fischbacher, 1998).

choice set of the standard dictator game and allows dictators not only to give something to the co-player, but also to take money from him. He observes that this modification of the experimental setting significantly reduces the average gift made by dictators. List argues that this change of behavior is due to the fact that the new setting puts much less weight on *giving* than the standard dictator game and, thus, reduces the influence of social norms that demand to give (i.e., dictators can demonstrate that they are not purely selfish by *not taking everything*). Our paper supports List's conjecture by providing further evidence on how the perception of the experimental situation alters subjects' 'give' and 'take' behavior.

Our research is inspired by the studies conducted by Fisman et al. and Andreoni and Miller. Similar to these studies we focus on preferences regarding the distribution of payoffs and, consequently, exclude any strategic interaction between subjects. The paper deviates from its precursors insofar as we investigate not only whether individual choices are consistent with some kind of other-regarding preferences, but also whether this behavior is stable over time. For this purpose we, first, confront subjects in a within-subject design with different variants of modified dictator games and sequential prisoner's dilemma games in order to check whether they behave *consistently* within and across different classes of games.[2] Second, we repeat this experiment three times with the same subjects (but different partners) within three months.[3] This allows us to investigate whether subjects' (non-strategic) behavior is *stable* over time.[4] Note that this repeated experiment setting is different to a game that is repeated in one session. While in the latter setting the repetitions are part of *one single* experimental situation, our design *repeatedly* puts subjects in the same experimental situation.

Some of the above-mentioned theories of other-regarding behavior make very specific assumptions about the functional form and the parameters of the utility function. For example, Fehr and Schmidt as well as Charness and Rabin use linear specifications with some very particular prop-

[2] Within and across game consistency was also investigated by Blanco, Engelmann, and Normann (2006). In contrast to our paper, they solely focus on the theory proposed by Fehr and Schmidt (1999) and do not aim to analyze the stability of individual preferences over time. Fischbacher and Gächter (2006) use a within-subject design to analyze individual preferences in two subsequently played public good games.

[3] In the following, we call these repetitions wave one, wave two, and wave three.

[4] It should be mentioned that, in contrast to the experiments run by Andreoni and Miller (2002) and by Fisman et al. (2007), our subjects play the games in only one role, i.e. they are either dictators or recipients. This design has the advantage that the dictators' behavior is not influenced by possible expectations about the amount they receive from other subjects.

erties. We do not regard it a fair test of these theories to estimate the parameters and to investigate whether subjects always exhibit the same parameters when they play different games. Therefore, we transform these theories into very general versions which capture the fundamental ideas underlying each theory, but do not make use of particular parameter constellations. This allows us to test whether subjects in our experiments meet the minimal requirements of a theory that assumes that decisions are based on rational choices made under the assumption that subjects harbor some kind of other-regarding preferences.

Section 2 of our paper describes the modified dictator games and sequential prisoner's dilemma games in more detail. In section 3, we present the notions of consistency tested in the experiments. These notions are based on the standard approach of purely self-interested behavior and on two prominent specifications of other-regarding preferences, namely inequality aversion (see, e.g., Fehr and Schmidt 1999, Bolton and Ockenfels 2000, and Charness and Rabin 2002), and altruism (see, e.g., Andreoni and Miller 2002). The experimental design is included in section 4 and the findings are discussed in section 5. Section 6 summarizes our results and concludes.

Our observations cast some doubt on the existence of stable and consistent other-regarding decision-making. In particular, we could not find any truly other-regarding behavior which is stable over time. While there are many subjects in the first wave who exhibit some kind of concern for others, this behavior quickly changes over time into more selfishness. Stable behavior is displayed only by those who behave strictly selfishly.

2 Games

In order to investigate individual preferences, we employ two types of games: modified dictator games as introduced by Andreoni and Miller (2002) and sequential prisoner's dilemma games first employed by Clark and Sefton (2001). The games are described in detail below.

2.1 Modified dictator games

Our dictator games differ from standard dictator games in one important aspect: Dictators distribute money between themselves and the recipients; the amount to be distributed, however, is not fixed, but varies systematically. In each game, the dictator has to choose between eleven different distributions of payoffs to himself, π_A, and to the recipient, π_B. There are two different types of dictator games used in the experiment: 'take' games and 'give' games.[5]

Take games

In each of the four take games, starting from the initial endowment (π_A^E, π_B^E) = (500, 500), player A (the dictator) can reduce player B's (the recipient) payoff by $d\pi_B$ in order to increase his own payoff by $d\pi_A$ at a constant relative price of $p_A = |d\pi_B / d\pi_A|$, such that $\pi_A = 500 + (1 / p_A) (500 - \pi_B)$. This budget constraint can be re-formulated as $\pi_B = 500 + p_A (500 - \pi_A)$. Accordingly, the 'budget line' has a slope of $d\pi_B / d\pi_A = - p_A$. The four games only differ with respect to this slope: In the first game, T1, we have $p_A = p_A^{T1} = 1/2$; in the remaining games, the values are $p_A^{T2} = 2/3$, $p_A^{T3} = 1$, and $p_A^{T4} = 2$, respectively. Except for the equal payoff distribution, all possible options in the take games are chosen in such a way that player A is assured of a higher payoff than player B, i.e., $\pi_A > \pi_B$. The experimental set-up for the four games is illustrated in Table 1.

Game	π	1	2	3	4	5	6	7	8	9	10	11
T1	π_A π_B	500, 500	600, 450	700, 400	800, 350	900, 300	1000, 250	1100, 200	1200, 150	1300, 100	1400, 50	1500, 0
T2	π_A π_B	500, 500	575, 450	650, 400	725, 350	800, 300	875, 250	950, 200	1025, 150	1100, 100	1175, 50	1250, 0
T3	π_A π_B	500, 500	550, 450	600, 400	650, 350	700, 300	750, 250	800, 200	850, 150	900, 100	950, 50	1000, 0
T4	π_A π_B	500, 500	525, 450	550, 400	575, 350	600, 300	625, 250	650, 200	675, 150	700, 100	725, 50	750, 0

Table 1: Payoffs in the four take games.

Give games

In each of the four give games, starting from the initial endowment (π_A^E, π_B^E) = (500, 500), player A (the dictator) can increase player B's (the recipient) payoff by $d\pi_B$ at a personal cost of $d\pi_A$ at a constant relative price of $p_A = |d\pi_B / d\pi_A|$, such that $\pi_A = 500 + (1 / p_A)(500 - \pi_B)$. Again, this budget constraint can be re-formulated as $\pi_B = 500 + p_A (500 - \pi_A)$. Accordingly, the 'budget line' has a slope of $d\pi_B / d\pi_A = - p_A$. The four games only differ with respect to this slope: In the

[5] Note that List (2007) and Bardsley (2008) use a somewhat related modification of dictator games. In their games, the subjects' action set is extended in a way that allows them to either give money to or take money from the recipient. They find that this variation significantly decreases the number of subjects giving a positive amount.

first game, G1, we have $p_A = p_A^{G1} = 2$; in the remaining games, the values are $p_A^{G2} = 3/2$, $p_A^{G3} = 1$, and $p_A^{G4} = 1/2$, respectively. Choices in the give games (except for the equal payoff distribution) grant a higher payoff to player B than to player A, i.e., $\pi_A < \pi_B$. The experimental set-up is illustrated in Table 2.

Game	π	1	2	3	4	5	6	7	8	9	10	11
G1	π_A π_B	500, 500	450, 600	400, 700	350, 800	300, 900	250, 1000	200, 1100	150, 1200	100, 1300	50, 1400	0, 1500
G2	π_A π_B	500, 500	450, 575	400, 650	350, 725	300, 800	250, 875	200, 950	150, 1025	100, 1100	50, 1175	0, 1250
G3	π_A π_B	500, 500	450, 550	400, 600	350, 650	300, 700	250, 750	200, 800	150, 850	100, 900	50, 950	0, 1000
G4	π_A π_B	500, 500	450, 525	400, 550	350, 575	300, 600	250, 625	200, 650	150, 675	100, 700	50, 725	0, 750

Table 2: Payoffs in the four give games.

2.2 Sequential prisoner's dilemma games

The payoffs in the two sequential prisoner's dilemma (PD) games are given in Figure 1. In both games, the decisions of player A (the second mover) are elicited using the strategy method[6], i.e. player A has to respond to each of the two actions feasible for player B (the first mover).

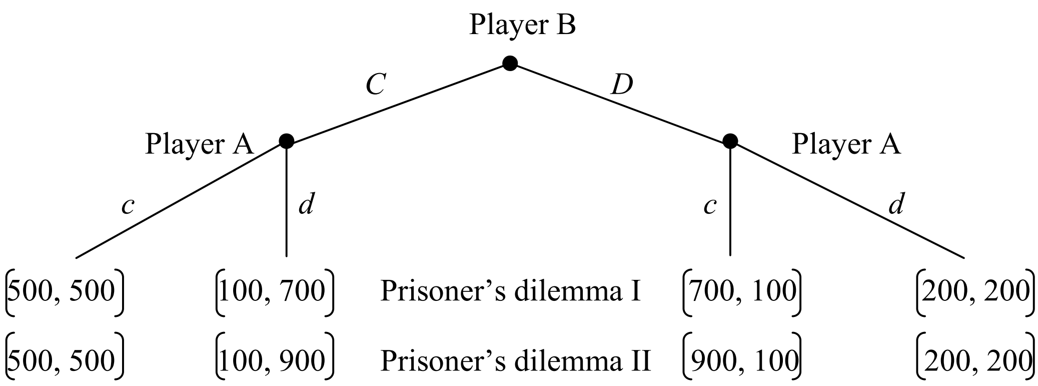

Figure 1: Payoffs in the two prisoner's dilemma games [π_B, π_A]

[6] We make use of the strategy method because this allows us to exclude any feedback about the opponent's behavior and, thus, eliminates all strategic considerations. As Brosig et al. (2003) have shown, applying this method might possibly influence behavior. Note, however, that even if this influence exists in our experiment, it is the same for all of our games, since all of them are played in a "cold" mode.

Note that the PD-subgames played by A can be interpreted as mini take and give games, where the take games are the ones following player B's *C*-move and the give games are the ones following player B's *D*-move.

2.2.1 Mini take games

The mini take game in PD I leaves player A the choice between (π_A^E, π_B^E) = (500, 500) and (π_A, π_B) = (700, 100). This creates a relative price of p_A = 2, such that the slope of the respective budget line is dπ_B / dπ_A = -2 (similar to game T_4).

In the mini take game of PD II, player A chooses between (π_A^E, π_B^E) = (500, 500) and (π_A, π_B) = (900, 100). Accordingly, we have p_A = 1. The slope of the respective budget line is dπ_B / dπ_A = -1 (similar to game T_3).

2.2.2 Mini give games

The mini give game in PD I gives player A the opportunity to choose between (π_A^E, π_B^E) = (200, 200) and (π_A, π_B) = (100, 700). This creates a relative price of p_A = 5, such that the slope of the respective budget line is dπ_B / dπ_A = -5.

In the mini give game of PD II, player A can decide between (π_A^E, π_B^E) = (200, 200) and (π_A, π_B) = (100, 900). Accordingly, we have p_A = 7. The slope of the budget line is dπ_B / dπ_A = -7.

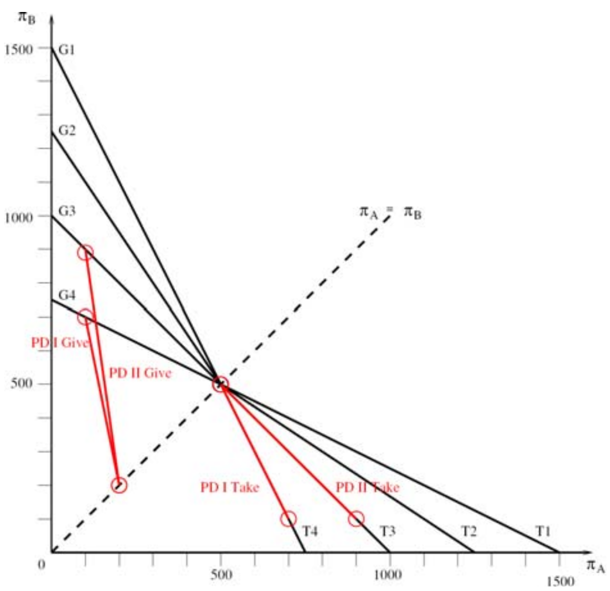

Figure 2: Budget lines in modified dictator games and prisoner's dilemma games.

3 Concepts of consistency

In our study, we investigate consistency with regard to the decisions made by players A. The concepts are based on three notions of preferences: **s**elfish preferences ('S-consistency'), **a**ltruistic preferences ('A-consistency'), and **i**nequality aversion preferences ('I-consistency'). It should be noted that the different concepts of consistency presented below are not mutually exclusive. There are many modes of behavior that can be consistently explained by more than just one of these concepts.

3.1 Consistency according to selfish preferences

A player A with selfish preferences cares solely about his own payoff π_A. Assuming that player A derives positive utility from π_A, i.e.

$$U_A^S = u(\pi_A) \quad \text{with} \quad \frac{\partial u(\cdot)}{\partial \pi_A} > 0,$$

then a selfish individual will always maximize his own payoff. The implications for the definition of consistency in our games are straightforward. An S-consistent player A will always take the maximum possible amount from player B in the take games (i.e., will always choose option 11), will always transfer the minimum possible amount to player B in the give games (i.e., will always choose option 1), and will always choose *d* in PD games.

All types of other-regarding preferences analyzed in this paper will include selfish preferences as a special case.

3.2 Consistency according to altruistic preferences

Our second concept of consistency is in the spirit of that introduced by Andreoni and Miller (2002). According to this concept, one's own and the other's payoff are considered as 'normal goods'. Assuming that player A derives non-negative utility from her own payoff and from player B's payoff and ruling out the case of simultaneous indifference with regard to both, the utility function can be written as follows:

$$U_A^A = u(\pi_A, \pi_B) \quad \text{with} \quad \frac{\partial u(\cdot)}{\partial \pi_A} \geq 0, \frac{\partial u(\cdot)}{\partial \pi_B} \geq 0 \quad \text{and} \quad \exists \frac{\partial u(\cdot)}{\partial \pi_i} > 0 \text{ for } i \in \{A, B\}.$$

Using the additional assumption that neither π_A nor π_B are inferior, we conclude that 'demand' for π_A and π_B must each be falling in its own respective price.

Dictator games

Since π_A and π_B are normal goods, optimum demand for π_B, π_B^*, should not decrease with the relative price p_A, i.e. $\partial\pi_B^*/\partial p_A \geq 0$. Consequently, in the take games, the amount taken from player B, $\tau = 500 - \pi_B^*$, should not increase with p_A, i.e. $\partial(500-\pi_B^*)/\partial p_A \leq 0$. Given that in the four take games $p_A^{T1} < p_A^{T2} < p_A^{T3} < p_A^{T4}$, our definition of consistency in these games is:

A-CONSISTENCY IN TAKE GAMES: An A-consistent player A in a take game will not decrease π_B (will not take more from player B) as the relative price p_A of his own payoff increases, i.e.

$$\tau_{T1} \geq \tau_{T2} \geq \tau_{T3} \geq \tau_{T4}.$$

Accordingly, in the give games, the amount given to player B, $\gamma = \pi_B^* - 500$, should not fall with p_A, i.e. $\partial(\pi_B^* - 500)/\partial p_A \geq 0$. Given that in the four give games $p_A^{G1} > p_A^{G2} > p_A^{G3} > p_A^{G4}$, our definition of consistency in the give games is:

A-CONSISTENCY IN GIVE GAMES: An A-consistent player A in a give game will not decrease π_B (will not give less to player B) as the relative price p_A of his own payoff increases, i.e.

$$\gamma_{G1} \geq \gamma_{G2} \geq \gamma_{G3} \geq \gamma_{G4}.$$

PD games

For players A with altruistic preferences, the definition of consistency in the prisoner's dilemma games is straightforward:

A-CONSISTENCY IN PD GAMES: An A-consistent player A who is following a *C* choice of player B (mini take games), should choose always *d* or always *c* in both PD games or should choose *c* in PD I and *d* in PD II.

An A-consistent player A who is following a *D* choice of player B (mini give games), should choose always *d* or always *c* in both PD games or should choose *d* in PD I and *c* in PD II.

3.3 Consistency according to inequality aversion preferences

Another way of modeling other-regarding preferences takes into account notions of inequality aversion. Preferences of this type have been introduced by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000); Charness and Rabin (2002) also consider inequality aversion, but additionally include reciprocity concerns in their preference model. Note that, in our games, considering reciprocity along the lines of Charness and Rabin does not change the predictions for individual behavior.[7] In the following, we present a more general notion of inequality aversion which is inspired by the three papers above.

With preferences including inequity aversion, a player gains utility from his own payoff and loses utility from a difference between his own and the other's payoff. The utility function can be written as follows

$$U_A^I = u(\pi_A, \Delta) \text{ with } \Delta = \begin{cases} \Delta_a = \pi_A - \pi_B & \text{for } \pi_A - \pi_B \geq 0 \quad (1) \\ \Delta_d = \pi_B - \pi_A & \text{for } \pi_B - \pi_A > 0 \quad (2) \end{cases},$$

were strict inequality aversion implies that $\frac{\partial u(\pi_A, \Delta)}{\partial \Delta} < 0$.

Consequently, in the first case (1) it holds that $\frac{\partial u(\pi_A, \Delta)}{\partial \pi_B} = \frac{\partial u(\pi_A, \Delta)}{\partial \Delta_a} \cdot \frac{\partial \Delta_a}{\partial \pi_B} > 0$.

(1a) If $\frac{\partial u(\pi_A, \overline{\Delta})}{\partial \pi_A} \geq -\frac{\partial u(\overline{\pi}_A, \Delta)}{\partial \Delta_a} \cdot \frac{\partial \Delta_a}{\partial \pi_A}$,[8] then $\frac{\partial u(\pi_A, \Delta)}{\partial \pi_A} = \frac{\partial u(\pi_A, \overline{\Delta})}{\partial \pi_A} + \frac{\partial u(\overline{\pi}_A, \Delta)}{\partial \Delta_a} \cdot \frac{\partial \Delta_a}{\partial \pi_A} \geq 0$ and the marginal rate of substitution (MRS) is non-positive, i.e.

$$\text{MRS} = -\frac{d\pi_B}{d\pi_A} = -\frac{\partial u(\pi_A, \Delta)/\partial \pi_A}{\partial u(\pi_A, \Delta)/\partial \pi_B} \leq 0.$$

(1b) If $\frac{\partial u(\pi_A, \overline{\Delta})}{\partial \pi_A} < -\frac{\partial u(\overline{\pi}_A, \Delta)}{\partial \Delta_a} \cdot \frac{\partial \Delta_a}{\partial \pi_A}$, then $\frac{\partial u(\pi_A, \Delta)}{\partial \pi_A} = \frac{\partial u(\pi_A, \overline{\Delta})}{\partial \pi_A} + \frac{\partial u(\overline{\pi}_A, \Delta)}{\partial \Delta_a} \cdot \frac{\partial \Delta_a}{\partial \pi_A} < 0$ and the MRS is positive, i.e. $\text{MRS} = -\frac{d\pi_B}{d\pi_A} = -\frac{\partial u(\pi_A, \Delta)/\partial \pi_A}{\partial u(\pi_A, \Delta)/\partial \pi_B} > 0$.

In the second case (2) it holds that $\frac{\partial u(\pi_A, \Delta)}{\partial \pi_B} = \frac{\partial u(\pi_A, \Delta)}{\partial \Delta_d} \cdot \frac{\partial \Delta_d}{\partial \pi_B} < 0$ and that

$\frac{\partial u(\pi_A, \Delta)}{\partial \pi_A} = \frac{\partial u(\pi_A, \overline{\Delta})}{\partial \pi_A} + \frac{\partial u(\overline{\pi}_A, \Delta)}{\partial \Delta_a} \cdot \frac{\partial \Delta_a}{\partial \pi_A} \geq 0$. Consequently, the MRS is non-negative, i.e.

$$\text{MRS} = -\frac{d\pi_B}{d\pi_A} = -\frac{\partial u(\pi_A, \Delta)/\partial \pi_A}{\partial u(\pi_A, \Delta)/\partial \pi_B} \geq 0.$$

---

[7] In particular, the specific concepts by Charness and Rabin and by Fehr and Schmidt would lead to the same results (see Appendix A for a proof).

[8] This case reflects the assumption made by Fehr and Schmidt (1999, p. 824) that the direct marginal utility of π_A always exceeds the indirect marginal loss caused by the greater payoff difference.

Dictator games

Following our general notion of inequality aversion, in the give games, case (2) applies and the MRS is always larger than $-p_A$ (i.e., the slope of the budget line). Consequently, an optimum choice will always be represented by a corner solution. Given that player A profits from a higher π_A and from a lower π_B, an I-consistent player A should keep everything to himself:

I-CONSISTENCY IN GIVE GAMES: An I-consistent player A in a give game will transfer no money to player B, i.e. $\gamma_{G1} = \gamma_{G2} = \gamma_{G3} = \gamma_{G4} = 0$.

Things are slightly more complicated if player A has a higher payoff than player B ($\pi_A \geq \pi_B$), as is the case in our take games. Changes in π_A will have two opposite effects. On the one hand, a rise in π_A increases utility. On the other hand, this also increases the inequality between π_A and π_B, which, in turn, decreases utility. Given behavior is best described by (1a), then three different notions of I-consistency in the take games apply. First, if the MRS is greater than $-p_A^{T1}$, then player A should always take nothing. Second, if the MRS is less than $-p_A^{T4}$, then player A should always take everything. Third, if the MRS is less than $-p_A^{T1}$, but greater than $-p_A^{T4}$, then player A should not take more from player B as the relative price p_A of his own payoff increases.[9] Given behavior is best described by (1b), then I-consistency implies that player A always takes nothing since the MRS is greater than $-p_A^{T1}$.

I-CONSISTENCY IN TAKE GAMES: An I-consistent player A in a take game will not take more from player B as the relative price p_A of his own payoff increases, i.e. $\tau_{T1} \geq \tau_{T2} \geq \tau_{T3} \geq \tau_{T4}$.

PD games

Given the previous remarks, for players A with inequality aversion preferences the definition of consistency in the prisoner's dilemma games is as follows:

I-CONSISTENCY IN PD GAMES: An I-consistent player A in both PD games who is following a *D* choice of player B (mini give games), should choose *d*. An I-consistent player A who is following a *C* choice of player B (mini take games), should choose always *d* or always *c* in both PD games or should choose *c* in PD I and *d* in PD II.[10]

[9] Note that these three notions require that preferences are homothetic.
[10] All three notions of I-consistency in the mini take games result when (1b) holds. Otherwise, I-consistency implies that players A always choose *c* in both PD games.

Our definitions of consistency obtained by applying selfish preferences, altruistic preferences, and inequality aversion preferences are summarized in Table 3. Note that, for our games, I-consistency is a special case of A-consistency, and S-consistency is a special case of I-consistency.

<table>
<tr><th></th><th>Give games</th><th>Take games</th><th>PD I</th><th>PD II</th></tr>
<tr><td rowspan="2">S-consistency</td><td rowspan="2">$\gamma_{G1} = \gamma_{G2} = \gamma_{G3} = \gamma_{G4} = 0$</td><td rowspan="2">$\tau_{T1} = \tau_{T2} = \tau_{T3} = \tau_{T4} = 500$</td><td colspan="2">always d/D</td></tr>
<tr><td colspan="2">always d/C</td></tr>
<tr><td rowspan="2">I-consistency</td><td rowspan="2">$\gamma_{G1} = \gamma_{G2} = \gamma_{G3} = \gamma_{G4} = 0$</td><td rowspan="2">$\tau_{T1} \geq \tau_{T2} \geq \tau_{T3} \geq \tau_{T4}$</td><td colspan="2">always d/D</td></tr>
<tr><td colspan="2">always d/C,
always c/C,
c/C in PD I and d/C in PD II</td></tr>
<tr><td rowspan="2">A-consistency</td><td rowspan="2">$\gamma_{G1} \geq \gamma_{G2} \geq \gamma_{G3} \geq \gamma_{G4}$</td><td rowspan="2">$\tau_{T1} \geq \tau_{T2} \geq \tau_{T3} \geq \tau_{T4}$</td><td colspan="2">always d/D
always c/D,
d/D in PD I and c/D in PD II</td></tr>
<tr><td colspan="2">always d/C,
always c/C,
c/C in PD I and d/C in PD II</td></tr>
</table>

Table 3: Definitions of consistency.

3.4 Consistency across games

S-consistency, A-consistency, and I-consistency in each of the ten games implies S-consistency, A-consistency, and I-consistency, respectively, across games.

4 Experimental design

The ten games were played with the same players A over two sessions which were conducted within one week. Each of the two sessions was run with four groups of subjects, each consisting of 10 players A and 10 players B. In session 1, subjects participated in the four take games and in prisoner's dilemma I. In session 2, subjects participated in the four give games and in prisoner's dilemma II. The sequence of play is illustrated in Table 4. The two sessions were repeated twice, once after four weeks (wave 2) and once after another four weeks (wave 3). Conducting the 2 x 3 sessions using a within-subject design for players A allows us to investigate the stability and consistency of their preferences. When signing up for the first two sessions, players A were told that they would possibly have to take part in more than one experiment, but were left ignorant about the number of experiments they would have to participate in. Players B were newly recruited for each session and players A were informed about this. This design feature should eliminate strategic considerations that could influence the stability of behavior.

	1st game	2nd game	3rd game	4th game	5th game
Session 1	T2	T4	PD I	T1	T3
Session 2	G3	G1	PD II	G4	G2

Table 4: Sequence of play.

At the beginning of each session in waves 1 and 2, subjects were told that they had to make decisions, but were left ignorant about the structure and number of games to be played.[11] Wave 3 differed from the previous two waves in that subjects were informed about the five games immediately at the beginning of each session.[12] In all sessions, we employed a perfect random matching design, i.e. players A were matched with different players B, and subjects were informed accordingly. In addition, subjects were told that they would receive no feedback about their partner's and others' decisions during the experiment. The purpose of this design was to prevent players A from "learning" about their opponents' behavior. Nevertheless, players A could possibly compute their opponents' (past) behavior in PD-games from their total profits and could condition their behavior in the next games, played with new opponents, on their former opponents' choices. However, such behavior is not observed in our data set. At the end of each session, subjects were paid off the total profit they had made in the five games at an exchange rate of 150 Lab-Cents = 100 Eurocents. The payment was conducted anonymously employing a double-blind procedure.

The computerized experiment was run with a total of 270[13] students using Fischbacher's (2007) z-Tree software tool. Average payoffs per session were €15.03, with a minimum of €0.67 and a maximum of €34.67. There was no show-up fee. No session lasted longer than 30 minutes.

[11] Instructions are included in Appendix D.
[12] This was done in order to amplify subjects' experience in a way that mimics the influence of repetitions of the experiment. Possible effects on subjects' behavior are discussed in section 5.1. In all sessions, subjects had to submit their choices one after the other.
[13] In wave 1, there were 40 (40) players A (B) in both sessions. Due to no-shows, in wave 2 there were 39 (39) players A (B) in both sessions, and in wave 3 there were 37 (37) players A (B) in session 1 and 35 (35) players A (B) in session 2.

5 Results

In order to report on our large set of data (decisions made by players A in three different types of games – take games, give games, and PD games – which are played in different variants in three waves over time)[14], the results section is structured in the following way. We first look at the aggregate data level, which is the focus of most experimental studies. After that, we take a closer inspection of *individual* behavior. This essentially means two things. First, we analyze the *consistency* of individual behavior, trying to find out to what extent the concepts of consistency introduced earlier in this paper can account for individual behavior observed within and across the three classes of games in each of the three waves. Second, we investigate each individual's *stability* of behavior, describing whether, and, if so, how, individual behavior changes over time.

Looking at aggregate behavior, we find that, in the first wave, the three models assuming some form of other-regarding behavior outperform the standard theory of pure self-interest. Similar findings apply to individual behavior. In line with the observations made by Fisman et al., there are more subjects who consistently display other-regarding behavior (particularly A-consistent) than subjects who consistently display selfish behavior within and across the different classes of games. Over time, the frequency of consistent behavior increases. The proportion of consistent purely other-regarding behavior declines from wave to wave, however, while the proportion of consistent self-interested behavior increases. In the third wave, nearly all consistent decisions can be rationalized by pure self-interest. This last observation also dominates our findings concerning the stability of behavior. Only a few subjects exhibit stable behavior over time, all of them making selfish decisions. The following subsections present our findings in more detail.

5.1 Aggregate behavior

In order to get a first impression of what happens in the course of the experiment, we look at aggregate data obtained for each class of games in each of the three waves.[15]

---

[14] Our data are included in Appendix C.

[15] Analyzing wave 3, we have to take into account that the experimental design changed slightly from wave 2 to wave 3. In the last wave, subjects were informed about the sequence of games, while in the first two waves they were not. Changing the informational design in this direction, we believe that we can generate a 'time-lapse' effect, mimicking the experience-enhancing effect of repeating the experiment. Potentially this change is in favor of non-selfish behavior. Given a subject plans to 'give' some money to one of his opponents, he can choose the one single game within the session that suits him most to do this (maybe because giving to others is particularly 'cheap' in that

In the take games of the first wave, we observe that the average amounts taken away from players B are lower than 500. Moreover, the average amount taken, if anything, significantly decreases with a lower relative price for the payoff of players A (significance levels for all game comparisons are displayed in Table 5). That is, at the aggregate level, the observed behavior is in line with both theories of other-regarding behavior (inequality aversion and altruism).

	T1 vs. T2	T1 vs. T3	T1 vs. T4	T2 vs. T3	T2 vs. T4	T3 vs. T4
Wave 1	p = 0.288	p = 0.132	p = 0.071	p = 0.036	p = 0.023	p = 0.266
Wave 2	p = 0.438	p = 0.367	p = 0.008	p = 0.041	p = 0.000	p = 0.039

Table 5: Significance levels for take games (two-tailed exact Wilcoxon test).

The results for the second wave are similar, though the average amounts taken away by players A are significantly higher than in wave 1 ($p < 0.003$, two-tailed exact Wilcoxon test). In the third wave, players A take almost all money from players B in all four games. There are no longer any significant differences regarding the behavior of players A across the four games. These observations indicate that selfish behavior, which is not dominant in wave 1, already plays a major role in wave 2 and dominates in the last wave. The aggregate results from the dictator games are displayed in Figure 3.

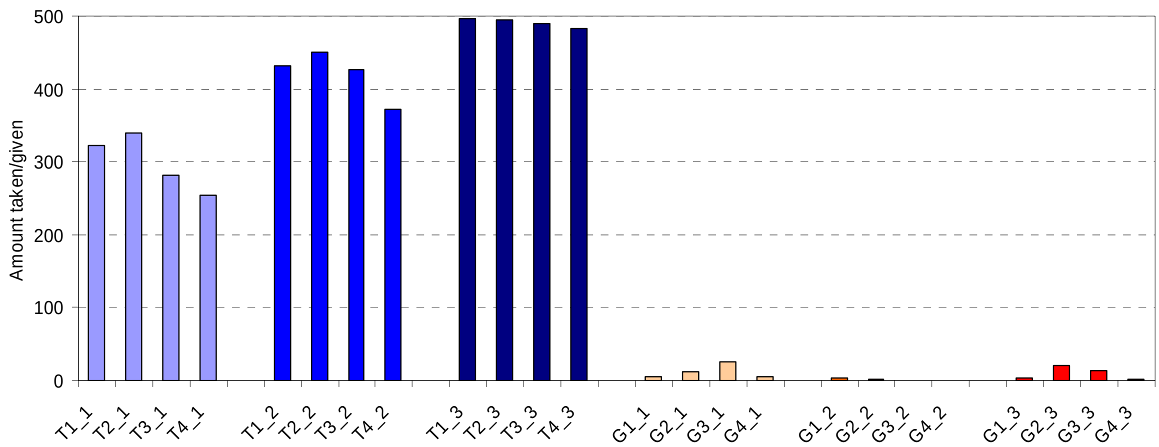

Figure 3: Average amount taken/given in the take games/give games in the three waves.

game, maybe because the game allows the subjects to give away a certain amount of money, or for other reasons). Calculating the total amount of money players A allocate to themselves or the total amount of money players A allocate to their opponents for all three waves separately, we do not observe such an increase in other-regarding behavior, however (see Figures B1 and B2 in Appendix B).

In the give games, things are quite different. On average, players A do not give significant amounts in any of the three waves. There are neither significant differences between games nor between waves. The fact that, in these games, giving creates an efficiency gain does not seem to be a driving force for average decisions. The aggregate data obtained in give games can be fully accounted for by the standard model of purely selfish behavior. Still, it is not contrary to the predictions made by the other three models.

A comparison of the average frequencies of *d*-moves between PD I and PD II reveals no significant differences within any one wave (see Figure 4). Similarly, there are no significant differences between the average response of players A to a *C*-move and their average response to a *D*-move in either of the two games and the three waves.[16] That is, the average player A does not make his move dependent on the behavior of his opponent. It should be noted in passing that this means that subjects do not show any kind of reciprocal behavior in our prisoner's dilemma games. This finding is at odds with the observations made by, e.g., Clark and Sefton (2001), but it is in line with inequality aversion and altruism

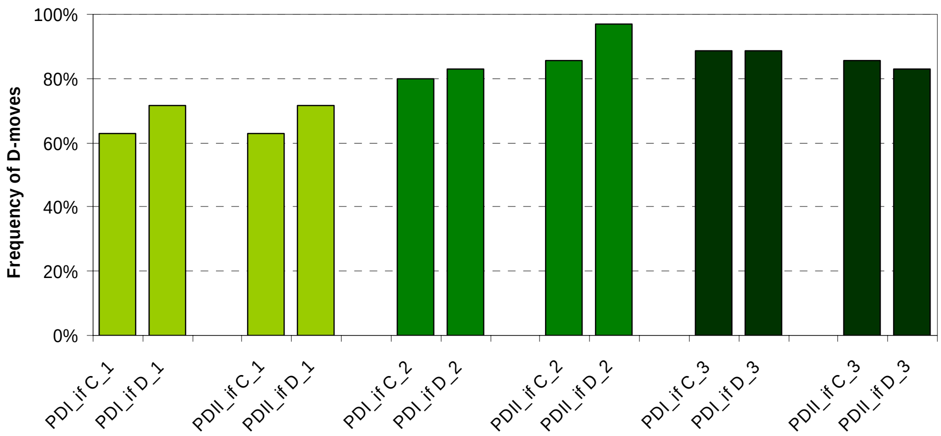

Figure 4: Average frequency of *d*-moves in the two PD-games in the three waves.

The average frequency of purely self-interested behavior is always higher than 60 percent in wave 1, and it increases in nearly all repetitions (except for PD II, where we observe a weakly significant decrease in *d*-moves from wave 2 to wave 3). From wave 2 on, the average frequency

[16] If not indicated otherwise, two-tailed exact McNemar tests are used. Differences are labeled as significant if $p < 0.05$ and are labeled as weakly significant if $0.05 \leq p < 0.1$.

of defection in no case drops below the 80 percent level – also in response to a cooperative move made by player B. Our results are summarized in observation 1:

Observation 1 (aggregate behavior)

In two of the three classes of games (take games and PD games), we observe that, in the first wave, subjects, on average, do not show strictly selfish behavior. Over the two repetitions, the fraction of purely self-interested decisions increases and, in the last wave, no more than 15 percent of decisions deviate from strict selfishness. In the give games, we observe selfish behavior right from the beginning.

5.2 Consistency of individual behavior

5.2.1 Consistency within games

In order to investigate whether subjects behave consistently within any one of the three classes of games, we use the three concepts of consistency introduced in section 3. Note that, as already mentioned, these concepts are not independent measures. Since selfishness (S-consistency) is a special case of I-consistency, and I-consistency is a special case of A-consistency in our games, the additional explanatory power of the latter concepts is limited to the cases where behavior is non-selfish or not I-consistent. Therefore, we should expect that most of the decisions are A-consistent and few are S-consistent. Figure 5 displays the results for all three waves:

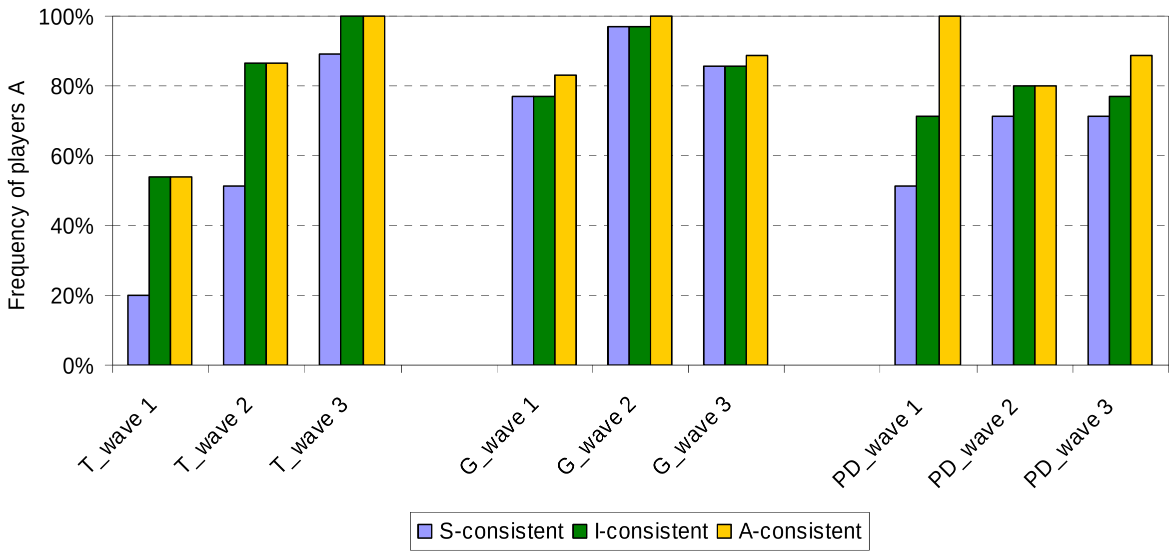

Figure 5: Frequency of consistent behavior in all games in the three waves.

For the take games, the concepts of I- and A-consistency are identical and explain about 50 percent of all individual moves in the first wave, whereas 18 percent of subjects display consistently selfish behavior. That is, roughly a little more than 30 percent of the observed decisions in the first wave are consistent if we take into account that subjects may harbor preferences as used in the I- and the A-concepts. In the second wave, we observe a highly significant increase in consistent behavior. Now, more than 80 percent of decisions are I- and A-consistent and this frequency is still significantly higher than the 49 percent of S-consistent decisions. The observation that the difference between both frequencies remains the same implies that the increase in consistency is mainly due to more consistent selfishness. In the third wave, the share of selfish behavior again increases significantly, this time to 89 percent. Only the decisions made by the 11 percent of subjects who do not take all the money from players B can be accounted for by I- and A-consistency only. Remarkably, we do not observe any individual move in wave 3 which cannot be characterized as consistent in the sense of one of the three concepts.

In the give games, two thirds of subjects behave selfishly right from the start. This share significantly increases from wave 1 to wave 2, but does not significantly change from wave 2 to wave 3. All three measures of consistency explain about 85 percent of observed individual behavior; there are no significant differences between the three measures in all three waves. Since those who decide to give something to player B choose only very small amounts, it seems to be adequate to characterize the overall behavior in the give games as "consistently selfish".

In the PD games, we find different patterns of behavior. All subjects behave in an A-consistent manner in the first wave. The reason for the good performance of A-consistency is the fact that all players A follow one of the four A-consistent patterns: they always defect, or always cooperate, or always play tit-for-tat, or always play 'inverted tit-for-tat', i.e. play *d*/*C* and *c*/*D*. Since I-consistent behavior is only in line with 'always *d*' and 'tit-for-tat', it could be observed in only 70 percent of all cases. 53 percent of subjects behave strictly S-consistent and always decide to defect. The behavioral patterns observed in wave 1 are summarized in Table 6.

	always *d*		always *c*		tit-for-tat		inverted tit-for-tat	
PD I	55.0%	(22/40)	15.0%	(6/40)	17.5%	(7/40)	12.5%	(5/40)
PD II	52.5%	(21/40)	15.0%	(6/40)	17.5%	(7/40)	15.0%	(6/40)
Both	**52.5%**	**(21/40)**	**15.0%**	**(6/40)**	**17.5%**	**(7/40)**	**12.5%**	**(5/40)**

Table 6: Behavior in the PD-games of wave 1.

In waves 2 and 3, the share of S-consistency increases to 72 percent and 71 percent, respectively. These shares are no longer significantly different from the shares of I- and A-consistency (for the latter this is only true in wave 2). The reason is that strategies that are not S-consistent ('tit-for-tat', 'inverted tit-for-tat', and 'always *c*') are hardly ever used any more. We summarize in observation 2:

Observation 2 (consistency within games):

The share of consistent moves within the three classes of games increases during the course of the experiment. This increase is almost always due to the fact that, over time, more and more subjects make consistently selfish decisions.

5.2.2 Consistency across games

The strongest test for the consistency concepts of individual behavior is the comparison of decisions made by a particular subject in different games. Figure 4 summarizes our findings concerning the consistency over all three classes of games.We employ the definition of across-game consistency introduced in section 3.4 in all three cases.

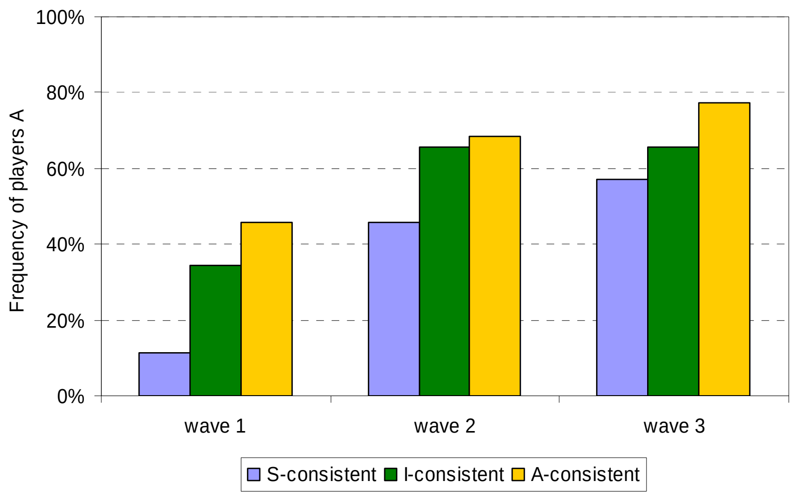

Figure 4: Consistency across games.

In the first wave, only 10 percent of all subjects exhibit consistently selfish behavior in all ten games. There are significantly more decisions that are I-consistent than S-consistent and significantly more that are A-consistent than I-consistent. A-consistency, the most general concept of consistency, accounts for the decisions made by 43 percent of subjects. The latter observation implies that, in wave 1, more than 50 percent of all subjects behave *in*consistently over the three classes of games. In wave 2, overall inconsistency decreases to 31 percent, which is largely due to a significant increase in the frequency of S-consistent behavior from wave 1 to wave 2. There is no longer a significant difference between the number of I- and A-consistent subjects. The shares of S-, I-, and A-consistent decisions do not significantly change from wave 2 to wave 3.

The rather low proportion of S-consistent behavior observed in wave 1 is due to the fact that subjects tend to make 'exceptions' to their otherwise selfish behavior while, in the last two waves, they consistently stick to their selfishness in all ten games. We summarize in observation 3:

Observation 3 (consistency across games):

While in the first wave selfishness is rarely observed across games, it dominates behavior in the second and the third waves. In the last wave, about three thirds of all decisions can be characterized as consistent across games.

5.3 Stability of individual behavior

Our concept of stability is rather simple. We denote individual behavior as *stable* if the subject always makes the same decision in the same game. As the question of stability is one of our major concerns, the results are presented in more detail by separately looking at the stability of individual behavior over waves 1 and 2, over waves 2 and 3, and over all three waves. Figures 5a-5c illustrate the stability of behavior observed in the three classes of games.

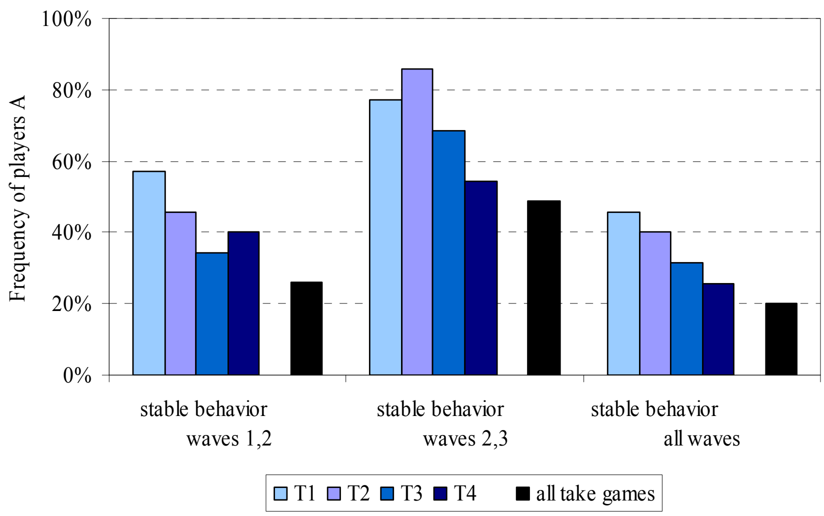

Figure 5a: Frequency of stable behavior in the take games.

Stability of individual choices is rather low in the four take games (Figure 5a). Only 23 percent of the subjects exhibit stable behavior from wave 1 to wave 2. This percentage significantly increases to 49 percent from wave 2 to wave 3. 20 percent of subjects display stable choices over all take games and over all three waves.

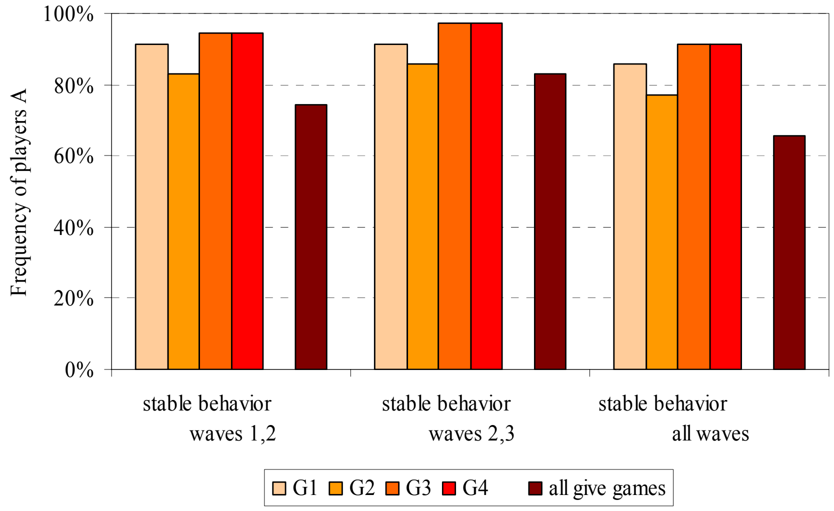

Figure 5b: Frequency of stable behavior in the give games.

We do not find a significant change in stable behavior over time in the give games (Figure 5b). The share of subjects making the same decisions in all four games over the first two waves is 72 percent while the share of subjects behaving in a stable manner over all three waves in this class of games is 66 percent.

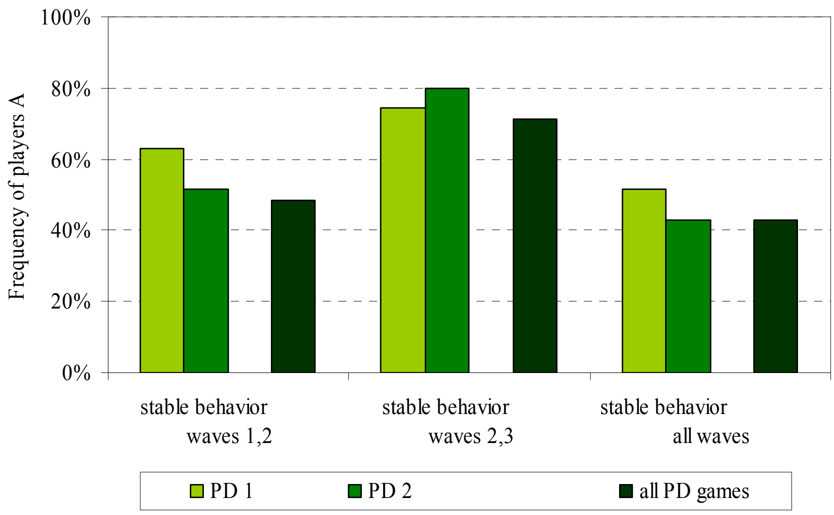

Figure 5c: Frequency of stable behavior in the PD games.

Similar to the take games, the frequency of stable behavior in the PD games (Figure 5c) significantly increases from the first two waves (49 percent) to the last two waves (71 percent). In total, 43 percent of subjects do not change their behavior over all three waves in the two PD games. Our findings are summarized in observation 4:

Observation 4 (stability):

Over all three waves, the frequency of stable behavior in the take games is rather low (20 percent) while in the PD games it is (significantly) higher (43 percent), and in the give games it is highest (66 percent).

5.4. Stability and consistency

It remains to combine the outcomes reported in the two preceding sections. How many subjects behave consistently *and* in a stable manner across games, and which of the concepts of consistency describes their behavior best?

	S-consistent	I-consistent, but not S-consistent	A-consistent, but not I-consistent
wave 1 to 2	5/39	1/39	0/39
wave 2 to 3	13/35	0/35	0/35
all waves	4/35	0/35	0/35

Table 8: Stable and consistent behavior

Whereas one of the six consistent decisions which are stable from wave 1 to wave 2 can be described by the concept of I-consistency alone, all of the I-consistent and all of the A-consistent choices which are stable over all three waves can be explained by the classical concept of S-consistency.[17] This finding directly leads to our fifth observation.

Observation 5 (stable consistency):

All modes of stable behavior over all games and all waves can be consistently explained by the assumption of selfish preferences.

[17] Accordingly, as observed in the previous section, there is a high frequency of stable behavior in those classes of games which reveal a high number of S-consistent decisions, i.e., the frequency of individual stability is (weakly) significantly higher in give games than in PD games and in take games and significantly higher in PD games than in take games.

6 Discussion and conclusion

During the last decade, considerable experimental evidence in favor of the existence of some kind of other-regarding preferences has been produced. The first wave of our experiment provides further support to these findings as most of the behavior observed in this wave cannot be explained by the assumption that subjects behave like rational egoistic payoff maximizers. In particular, the observations of aggregate and individual behavior in wave 1 leave room for explanations along the lines of theories assuming some kind of other-regarding preferences. For example, the results of our give games and of our take games in the first wave seem to largely confirm the assumption common for theories of inequality aversion that subjects are more willing to accept inequality when they are better off than their opponents than when they are less well off.

The major insight of our investigations is that, over the two repetitions of our experiment, subjects change their behavior tremendously. These changes have one unique direction which is common to all subjects: they behave more and more in a purely self-interested manner. In particular, stable behavior over time is observed only for those subjects who make strictly selfish decisions. It should be stressed that the major changes in individual behavior occur rather early, i.e., between the first two waves.

Given these findings, several more or less fundamental questions inevitably arise. The first line of questions seems to be quite obvious: Why is the observed behavior so unstable and why do subjects who do not initially exhibit self-interested behavior turn out to be *homines oeconomici* in the end? Two plausible (though speculative) explanations are at hand. First, it might be the case that subjects learn to be selfish in the sense that they find out that it does not hurt *not* to care about others. Therefore, in the third wave, they know that there is no internal punishment mechanism (bad feelings, bad conscience) at work when they take all the money, give nothing, and defect in the PD games. The second possible explanation is that subjects feel obliged to care about others, but that this obligation is finally fulfilled by forgoing strictly selfish behavior just once (independent of the fact that they are matched with new opponents in each of the three waves). Consequently, subjects in later waves might have the impression that they have done their duty and are in a position in which it is justified to care only about their own payoff. Both interpretations of behavior imply that subjects perceive the experimental situation in wave 2 and wave 3 in a different way than that in wave 1. Following List's (2007) conjecture, this different perception of the latter two waves might trigger the deactivation of social norms that are still at work in the first wave. In particular, the social norm that guided some subjects in wave 1 to leave a significant amount of money for the other player seems to lose its power in the last two waves.

This finally leads to a second line of questions relating to a more general, methodological point. Given our results, the question is: what is the relevant experimental evidence? Do we learn from our experiment that people behave selfishly or that they are not selfish in general? The answer to this question depends on the specific wave we look at and this leads to the question of what "true" experimental evidence is. Is it the behavior we observe when we invite subjects to the laboratory for the first time (which is the case in most of the experimental studies), or do we have to give subjects the chance to become familiar with the experimental situation? What is of greater importance, the behavior of the 'inexperienced' subjects in the first wave or the behavior of 'mature' subjects in the final wave? These methodological questions seem to be fundamental for experimental research and justify further investigations of the stability of laboratory observations.

7 References

Andreoni, J., Castillo, M., Petrie, R. (2003). “What do Bargainers’ Preferences Look Like? Exploring a Convex Ultimatum Game.” *American Economic Review* 93, 672-685.

Andreoni, J., Miller, J. H. (2002). “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism.” *Econometrica* 70, 737-753.

Bardsley, N. (2008). “Dictator Game Giving: Altruism or Artefact? A Note.” *Experimental Economics* 11, 122-133.

Blanco, M., Engelmann, D., Normann, H.-T. (2006). “A Within-Subject Analysis of Other-Regarding Preferences.” *Working Paper*.

Bolton, G. E., Ockenfels A. (2000). “ERC: A Theory of Equity, Reciprocity and Competition.” *American Economic Review* 90, 166-193.

Brosig, J., Weimann, J., Yang, C.-L. (2003). "The Hot Versus Cold Effect in a Simple Bargaining Experiment.” *Experimental Economics* 6, 2003, 75-90

Charness, G., Rabin, M. (2002). “Understanding Social Preferences with Simple Tests.” *Quarterly Journal of Economics* 117, 817-869.

Clark, K., Sefton, M., (2001) “The Sequential Prisoner's Dilemma: Evidence on Reciprocation”. *The Economic Journal* 111, 51–68.

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

Falk, A., Fischbacher, U. (2006) “A Theory of Reciprocity.” *Games and Economic Behavior* 54, 293-316.

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

Fischbacher, U. (2007). “z-Tree, Toolbox for Readymade Economic Experiments.” *Experimental Economics* 10, 171-178.

Fischbacher, U., Gächter, S. (2006). “Heterogeneous Social Preferences and the Dynamics of Free-Riding in Public Goods.” *CeDEx Discussion Paper No. 2006-01*, University of Nottingham.

Fisman, R., Kariv, S., Markovits, D. (2007). “Individual Preferences for Giving.” *American Economic Review* 97, 1858-1876.

Flood, M. M. (1952). “Some Experimental Games.” *Research Memorandum RM–789*, RAND Corporation.

Flood, M. M. (1958). “Some Experimental Games.” *Management Science* 5, 5-26.

Forsythe R., Horowitz J. L., Savin N. E., Sefton M. (1994). “Fairness in Simple Bargaining Experiments.” *Games and Economic Behavior* 6, 347-369.

Geanakoplos, J., Pearse, D., Stacchetti, E. (1989). “Psychological Games and Sequential Rationality”, *Games and Economic Behavior* 1, 60-80.

Isaac, R. M., McCue, K. F., Plott, C. R. (1985). “Public Goods Provision in an Experimental Environment.” *Journal of Public Economics* 26, 51-74.

Kahneman, D., Knetsch, J., Thaler, R. (1986). “Fairness and the Assumptions of Economics.” *Journal of Business* 59, S285-S300.

Kim, O., Walker, M. (1984). “The Free Rider Problem: Experimental Evidence.” *Public Choice* 43, 3-24.

Levine, D. (1998). “Modelling Altruism and Spitefulness in Experiments.”, *Review of Economic Dynamics* 1, 593-622.

List, J. A. (2007). “On the Interpretation of Giving in Dictator Games” *Journal of Political Economy* 115, 482-493.

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

Appendix A: Equivalence of the Fehr and Schmidt (1999) and the Charness and Rabin (2002) preference models in our games

For a two-person game with players A and B, Fehr and Schmidt (1999) specify individual A's preferences concerning his own payoff π_A and his opponent's payoff π_B:

$$U_A = u(\pi_A, \pi_B) = \begin{cases} \pi_\mathrm{A} - \beta(\pi_A - \pi_B) & \text{for } \pi_A \geq \pi_B \\ \pi_\mathrm{A} - \alpha(\pi_B - \pi_A) & \text{for } \pi_A < \pi_B \end{cases}.$$

The respective specification by Charness and Rabin (2002, p.822) reads

$$U_A = u(\pi_A, \pi_B) = \begin{cases} (\rho + \theta q)\pi_\mathrm{B} + (1 - \rho - \theta q)\pi_A & \text{for } \pi_A \geq \pi_B \\ (\sigma + \theta q)\pi_\mathrm{B} + (1 - \sigma - \theta q)\pi_A & \text{for } \pi_A < \pi_B \end{cases}$$

where q is an indicator variable signaling the presence of reciprocity. In case the opponent, player A, "misbehaves" (Charness and Rabin , 2002, p. 822), q becomes $q = -1$. As we designed our experiments in a way that avoids reciprocity, we can, for our paper, set $q \equiv 0$, which leads to the simple specification

$$U_A = u(\pi_A, \pi_B) = \begin{cases} \rho\pi_\mathrm{B} + (1 - \rho)\pi_A & \text{for } \pi_A \geq \pi_B \\ \sigma\pi_\mathrm{B} + (1 - \sigma)\pi_A & \text{for } \pi_A < \pi_B \end{cases}.$$

This can be re-written as

$$U_A = u(\pi_A, \pi_B) = \begin{cases} \pi_\mathrm{A} - \rho(\pi_A - \pi_B) & \text{for } \pi_A \geq \pi_B \\ \pi_\mathrm{A} + \sigma(\pi_B - \pi_A) & \text{for } \pi_A < \pi_B \end{cases}.$$

This form shows that, in our games, the specification by Fehr and Schmidt and the one by Charness and Rabin are equivalent for $\rho = \beta$ and $\sigma = -\alpha$.

Since sequential prisoner's dilemma games represent strategic interactions, reciprocity might play a role. According to Charness and Rabin, reciprocity only matters, however, if player B 'misbehaves', i.e. if player B chooses *D*. Following the *D*–move, player A can achieve both a higher π_A and a lower difference between π_A and π_B by choosing *d*. That is, utility increases even more when considering reciprocity along the lines of Charness and Rabin.

Appendix B: Figures

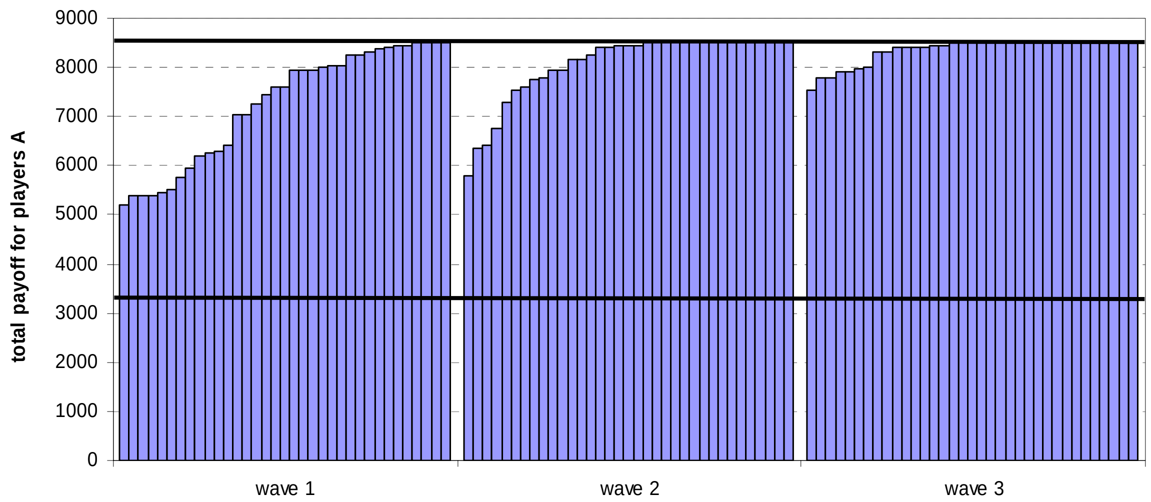

Figure B1: Total amount allocated by players A to themselves (includes all decisions by player A in the PD games; black lines indicate the maximum and minimum amounts).

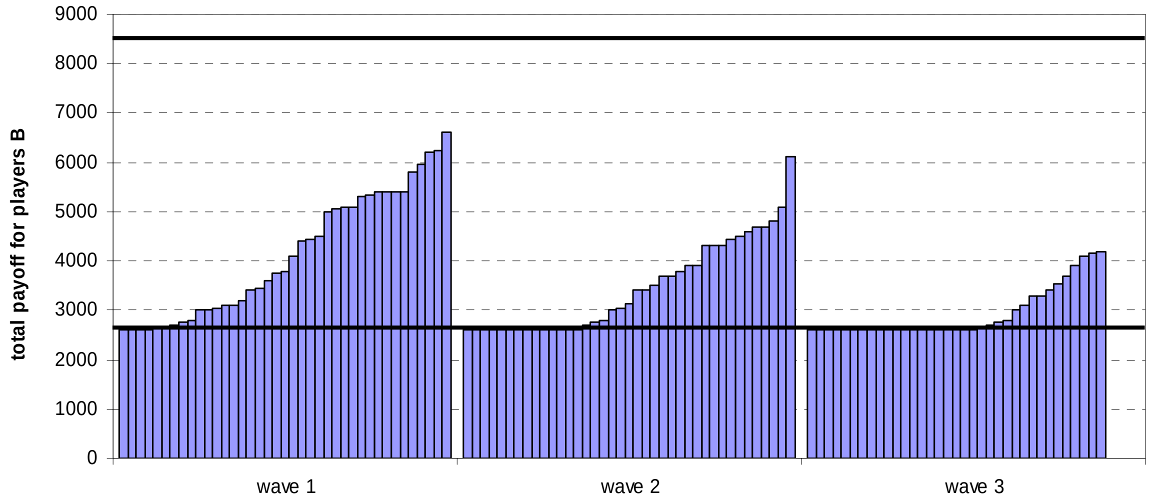

Figure B2: Total amount allocated by players A to player B (includes all decisions by player A in the PD games; black lines indicate the maximum and minimum amounts).

Appendix C: Data

Table C.1: Individual τ in Take Games

wave	**1**				**2**				**3**			
game	**T1**	**T2**	**T3**	**T4**	**T1**	**T2**	**T3**	**T4**	**T1**	**T2**	**T3**	**T4**
subject												
1	350	450	350	450	500	500	500	500	500	500	500	500
2	500	500	500	500	500	500	500	500	500	500	500	500
3	500	500	50	0	200	0	500	0	500	500	500	500
4	0	0	0	0	500	500	400	500				
5	0	0	0	400	500	500	500	500	500	500	500	500
6	0	400	50	0	500	500	500	0	500	500	500	500
7	0	0	50	0	0	500	0	0	500	500	500	500
8	450	400	350	200	500	500	500	500	500	500	500	500
9	0	200	50	500	500	500	500	0	500	500	500	500
10	200	250	200	200								
11	500	500	500	500	500	500	500	500	500	500	500	500
12	450	500	500	500	500	500	500	500	500	500	500	500
13	0	0	500	0	0	0	0	0				
14	450	450	450	450	450	450	450	450	500	500	450	450
15	100	150	100	50	400	400	400	400	500	500	500	500
16	100	50	100	150	200	500	200	500	500	500	500	500
17	500	500	500	500	500	500	500	500	500	500	500	500
18	0	200	0	0	500	500	500	500	500	500	500	500
19	500	500	450	450	500	500	500	500	500	500	500	500
20	0	0	0	0	0	0	0	0	500	500	500	500
21	250	200	400	500	500	500	450	450	500	500	500	500
22	500	500	500	500	500	500	500	500	500	500	500	500
23	500	450	500	500	500	500	250	300	500	500	500	500
24	500	450	400	350	500	500	500	400	500	500	500	500
25	500	500	500	500	500	500	500	500	500	500	500	500
26	0	0	0	0	500	500	450	400	500	500	500	500
27	500	500	250	400	500	500	500	500	500	500	500	500
28	500	500	200	0	500	500	500	500	500	500	500	500
29	500	300	300	500	500	500	500	500	500	500	500	500
30	500	500	500	500	500	500	500	500	500	500	500	500
31	200	250	0	0	300	250	150	0	350	350	300	200
32	500	500	500	0	500	500	500	400	500	500	500	500
33	0	100	50	50	0	150	100	50	500	500	500	500
34	0	0	0	0	500	500	450	200	500	500	500	500
35	200	300	300	0	450	450	400	350	500	450	450	400
36	500	500	500	500	500	500	500	500	500	500	500	500
37	500	500	400	450	500	500	500	350	500	500	500	400
38	500	500	500	0	500	500	500	500	500	500	500	500
39	400	400	200	0	500	500	500	500	500	500	500	500
40	500	300	400	0	500	500	500	500	500	500	450	450

Table C.2: Individual γ in Give Games

wave	1				2				3			
game	**G1**	**G2**	**G3**	**G4**	**G1**	**G2**	**G3**	**G4**	**G1**	**G2**	**G3**	**G4**
subject												
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	50	0	0	0	0	0	0	0	0	0	0
3	0	50	0	100	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0				
5	0	0	0	0	0	0	0	0	0	0	0	0
6	100	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0								
11	0	0	500	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0
13	500	0	0	500	0	0	0	0				
14	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0
21	0	200	0	0	0	0	0	0	0	500	450	0
22	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	400	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	50
28	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0				
30	0	0	0	0	0	0	0	0	0	0	0	0
31	200	100	0	0	0	0	0	0				
32	0	0	0	0	0	0	0	0	0	0	0	0
33	50	50	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	100	50	0	0	0	0	0	0
35	0	50	0	100	0	0	0	0	50	100	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	50	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	50	50	0	0

Table C.3: Individual Action Choices in PD games

wave	**1**				**2**				**3**			
game	**PD I**		**PD II**		**PD I**		**PD II**		**PD I**		**PD II**	
	if *C*	**if *D***	**if *C***	**if *D***	**if *C***	**if *D***	**if *C***	**if *D***	**if *C***	**if *D***	**if *C***	**if *D***
subject												
1	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*c*	*c*
2	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
3	*c*	*d*	*c*	*d*	*d*	*d*	*c*	*c*	*d*	*c*	*c*	*c*
4	*d*	*c*	*d*	*c*	*d*	*d*	*d*	*d*				
5	*c*	*c*	*c*	*c*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
6	*c*	*c*	*c*	*c*	*c*	*c*	*c*	*d*	*c*	*d*	*c*	*c*
7	*c*	*d*	*c*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
8	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
9	*d*	*c*	*d*	*c*	*d*	*c*	*d*	*d*	*d*	*d*	*d*	*c*
10	*d*	*d*	*d*	*c*								
11	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
12	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
13	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*				
14	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
15	*c*	*d*	*c*	*d*	*c*	*c*	*c*	*d*	*d*	*d*	*d*	*c*
16	*d*	*c*	*d*	*c*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
17	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
18	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
19	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
20	*c*	*c*	*c*	*c*	*c*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
21	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
22	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
23	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*c*	*c*	*c*	*d*
24	*d*	*c*	*d*	*c*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
25	*d*	*c*	*d*	*c*	*d*	*c*	*d*	*d*	*d*	*c*	*d*	*d*
26	*c*	*d*	*c*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
27	*c*	*d*	*c*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
28	*d*	*d*	*d*	*d*	*c*	*c*	*c*	*d*	*d*	*d*	*d*	*d*
29	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*		
30	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
31	*d*	*d*	*d*	*d*	*c*	*c*	*d*	*c*	*c*	*c*		
32	*c*	*c*	*c*	*c*	*c*	*c*	*c*	*d*	*c*	*d*	*c*	*d*
33	*c*	*d*	*c*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
34	*c*	*d*	*c*	*d*	*c*	*d*	*d*	*d*	*c*	*d*	*d*	*d*
35	*c*	*c*	*c*	*c*	*c*	*d*	*d*	*d*	*d*	*c*	*d*	*c*
36	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
37	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
38	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
39	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*
40	*c*	*c*	*c*	*c*	*d*	*d*	*d*	*d*	*d*	*d*	*d*	*d*

Appendix D: Instructions (*for players A; similar instructions were handed out to players B*)

Note

You are participating in an investigation of individual decision behavior. If you have any questions which are not answered by these instructions, please let us know. We will come to you and answer your questions.

During this experiment you will earn money. It depends on your decisions during the experiment how much money this will be. At the end of the experiment, the money will be paid to you in room C-214 when you display your ID card there.

Decision

During the experiment, you will have to make decisions at the computer. Before every decision, you will be given detailed instructions on the computer screen.

There is one other participant involved in each of your decision situations. This other participant will be newly allocated to you in each of your decision situations. We have made sure that you will interact with one and the same participant only once. No participant learns the identity of his allocated partners either during or after the experiment. Your decisions remain **anonymous**.

Please keep in mind that your decision situations are independent of one another, which means that none of your decisions has an influence on the other decisions.

[*Wave 3 only:*

In the appendix, you will find a list of the decision situations you will be confronted with during this experiment.]

Payoff

At the end of the experiment, we will compute your payoff in the laboratory. The exchange rate of laboratory cents to EURO cents is 150 Laboratory Cents = 100 EURO cent. The payoff to each of the participants will be put into an envelope, the envelope will be closed, and labeled with the respective ID number. The envelope will then be brought to room C-214 where a member of the staff who was not involved in the computation of the payoffs and who is sitting behind a blind will hand out your payoff when you display your ID card. This procedure makes sure that your decisions remain anonymous vis-a-vis the other participants and vis-a-vis the experimenter.

Please do not communicate with the other participants during the experiment. Moreover, we would ask you not to talk about the experiment to others in order to avoid influencing the behavior of potential future participants.

We thank you for your participation.

Decision situation 1:

Here you can determine the payoff to yourself and the payoff to your partner. Please choose one of the following combinations of payoffs:

choice:	○	You:	500	Partner:	500
	○	You:	450	Partner:	550
	○	You:	400	Partner:	600
	○	You:	350	Partner:	650
	○	You:	300	Partner:	700
	○	You:	250	Partner:	750
	○	You:	200	Partner:	800
	○	You:	150	Partner:	850
	○	You:	100	Partner:	900
	○	You:	50	Partner:	950
	○	You:	0	Partner:	1000

Decision situation 2:

Here you can determine the payoff to yourself and the payoff to your partner. Please choose one of the following combinations of payoffs:

choice:	○	You:	500	Partner:	500
	○	You:	450	Partner:	600
	○	You:	400	Partner:	700
	○	You:	350	Partner:	800
	○	You:	300	Partner:	900
	○	You:	250	Partner:	1000
	○	You:	200	Partner:	1100
	○	You:	150	Partner:	1200
	○	You:	100	Partner:	1300
	○	You:	50	Partner:	1400
	○	You:	0	Partner:	1500

Decision situation 3:

Please choose one of the "strategies" A or B for every possible choice of your partner. Your payoff depends on what you choose and what your partner chooses.

The following table displays your payoffs. If your partner chooses "A" and you choose "A if partner chooses A", your payoff is 500 (Your payoff is given by the second entry of the respective cell in the table given below.) If your partner chooses "A" and you choose "B if partner chooses A", your payoff is 900 and the payoff of your partner is 100. If your partner chooses "B" and you choose ""A if partner chooses B", your payoff is 100 and the payoff of your partner is 900. In the case that your partner chooses "B" and you choose "B if partner chooses B", your payoff is 200 and the payoff of your partner is also 200.

	you choose A	you choose B
Your partner chooses A	500 , 500	100 , 900
Your partner chooses B	900 , 100	200 , 200

Your choice, if partner chooses "A" ○ A
○ B

Your choice, if partner chooses "B" ○ A
○ B

Decision situation 4:

Here you can determine the payoff to yourself and the payoff to your partner. Please choose one of the following combinations of payoffs:

choice:
○ You: 500 Partner: 500
○ You: 450 Partner: 525
○ You: 400 Partner: 550
○ You: 350 Partner: 575
○ You: 300 Partner: 600
○ You: 250 Partner: 625
○ You: 200 Partner: 650
○ You: 150 Partner: 675
○ You: 100 Partner: 700
○ You: 50 Partner: 725
○ You: 0 Partner: 750

Decision situation 5:

Here you can determine the payoff to yourself and the payoff to your partner. Please choose one of the following combinations of payoffs:

choice:	○	You:	500	Partner:	500
	○	You:	450	Partner:	575
	○	You:	400	Partner:	850
	○	You:	350	Partner:	925
	○	You:	300	Partner:	1000
	○	You:	250	Partner:	1025
	○	You:	200	Partner:	1050
	○	You:	150	Partner:	1125
	○	You:	100	Partner:	1200
	○	You:	50	Partner:	1225
	○	You:	0	Partner:	1250