

A Within-Subject Analysis of Other-Regarding Preferences*

Mariana Blanco Dirk Engelmann Hans-Theo Normann

Royal Holloway College, University of London[†]

September 19, 2006

Abstract

We assess the predictive power of a model of other-regarding preferences, inequality aversion, using a within-subjects design. We run four different experiments (ultimatum game, dictator game, sequential prisoner's dilemma and public-good game) with the same sample of subjects. From the data we estimate parameters of aversion to disadvantageous and advantageous inequality. We then use these estimates to test several hypotheses across games. Our data show that results from within-subject tests can differ markedly from aggregate-level analysis. The inequality-aversion model has predictive power at the aggregate level but performs less well at the individual level.

JEL Classification numbers: C72, C91.

Keywords: behavioral economics, experimental economics, inequality aversion, other-regarding preferences

*We are grateful to Martin Dufwenberg, Ernst Fehr, Werner Güth, Wieland Müller, Nikos Nikiforakis, Avner Shaked, Juan Vargas and seminar and conference audiences at Ben-Gurion University, CERGE-EI Prague, Economic Science Association (ESA) Alessandria and Nottingham, European Economic Association (EEA) Vienna, Exeter University, Innsbruck University, Royal Economic Society (RES) Nottingham, and Royal Holloway College for helpful comments.

[†]Department of Economics, Royal Holloway College, University of London, Egham, Surrey TW20 0EX, UK, fax: +44 1784 439534, e-mail: *m.blanco@rhul.ac.uk*, *dirk.engelmann@rhul.ac.uk*, *hans.normann@rhul.ac.uk*.

1 Introduction

Currently, behavioral economists are putting a great deal of effort into developing models of other-regarding preferences. This literature takes as its starting point that data from, for example, the ultimatum game (Güth et al., 1982), the dictator game (Kahneman et al., 1986; Forsythe et al., 1994) and gift-exchange and trust games (Fehr et al., 1993; Berg et al., 1995) are by and large not compatible with the self-interested utility maximizing behavior of the traditional economic paradigm. The behavioral models attempt to explain these experimental results by relaxing the assumptions of the standard model and allowing for other-regarding motives.

Models of other-regarding preferences take into account the fact that participants' behavior is usually heterogenous (Fehr and Schmidt, 2006). A robust finding of various experimental studies is that not all participants behave in the same way. Not all subjects behave as the traditional model asserts they do, that is entirely selfishly, but neither do all subjects behave entirely, say, altruistically. The heterogeneity of the subjects' behavior suggests that models of other-regarding preferences should typically make predictions about how the distribution of behavior varies over different games.

This paper starts by noting that there are two ways in which predictions of this kind can be tested. To the best of our knowledge, the existing literature has relied almost exclusively on *aggregate-level* tests. However, the validity of the model can also be tested within subjects, that is, with *individual-level* analyses.¹ Consider an example with two games. In both games subjects may behave either “altruistically” or not, and some model

¹Aggregate-level analysis means that studies look for consistency of the *distribution* of choices across the experiments without considering whether *individual* subjects decide consistently with the theory across experiments. Aggregate-level analysis does not mean here that an investigation focusses on aggregate measures of the experimental data (such as means etc.).

implies that the condition for altruistic behavior according to a model parameter is the same in both games. An aggregate-level analysis can test the model by comparing the share of subjects who behave altruistically in the two games. If the share of these subjects is roughly the same in the two games, we would accept the model. The within-subjects test would involve checking whether individual subjects make altruistic (or non-altruistic) choices in both games. Noise in the data would of course mean that all subjects could not be expected to behave consistently across games, but, for individual-level validity of the model, the subjects who behave altruistically in the first game should have to be more likely than the other subjects to make the altruistic choice in the second game. In other words, accepting the model requires that behavior is correlated across games. The individual-level test obviously requires related-sample data whereas the existing literature has only used unrelated-sample data and, thus, has only made aggregate-level tests.

We believe that such individual-level data tests provide an interesting and useful additional test of other-regarding preferences theories, in particular when it comes to their behavioral validity. All other-regarding preferences theories that we are aware of are microeconomic models of decisions making, and explicitly state individual-level predictions. The motivation for the development of other-regarding preferences theories is to obtain a better description of behavior in experiments. Whether these accurately describe individual behavior is therefore a central issue. It is generally considered that the other-regarding preferences models predict aggregate outcomes well across several games. While this constitutes remarkable progress in the interpretation of recent experimental findings, we believe this success is incomplete without support from individual-level data. If a theory of individual decision making works at the aggregate level but the data show that individual behavior does not conform to the theory, we do not know what links the behavioral assumptions of the theory with its aggregate-level support. Put another way, given the choice between two comparable theories which work equally well at the aggregate level but only one of them

is also supported at the individual level, preference would surely be given to the theory that works at the individual level. One would simply feel more comfortable with a prediction from a model that consistently reflects individual behavior. We therefore believe a within-subject analysis of other-regarding preferences theories is both useful and warranted.

The lack of a comprehensive individual-level data study of other-regarding preference theories is indeed somewhat surprising. Two prominent papers argue in favor of it. Fehr and Schmidt (1999, p. 847) welcome this approach as “one of the most interesting tests of our theory”. Similarly, Andreoni et al. (2003, p. 683) argue that the comparison of aggregate and individual-level data “gives a new and interesting dimension to the analysis of experimental data”. What is more, individual-level experimental data can easily be produced. In fact, any related-sample design allows for the analysis of individual behavioral patterns. To test a theory of individual behavior with aggregate-level data is a plausible way of proceeding when individual-level data are not available or not reliable (for example, field data on voting). This, however, is not the case with the experiments that motivated the other-regarding preferences literature.

The main novelty of this paper is that it provides a first systematic² individual-level data test of a behavioral theory. The behavioral theory we analyze is a model of inequality aversion. This model was first proposed by Bolton (1991) and was refined by Fehr and

²Andreoni and Miller (2002) conduct a within-subjects analysis across several dictator games with different costs of giving. The objective of their study is to test whether subjects are consistent with the generalized axiom of revealed preferences and they find that most subjects are consistent. Andreoni et al. (2003) also conduct an individual-level comparison in their paper. However, they analyze only one game (a modified ultimatum game where subjects play both the proposer and the responder roles), and the individual-level comparison of decisions is arguably not the central part of the analysis (see also Section 5 below). Charness and Rabin (2002) use a related-sample design where subjects play two or four different games. They explicitly mention that they do not conduct an individual-level data analysis.

Schmidt (1999) and Bolton and Ockenfels (2000). Basically, inequality aversion stipulates that individuals do not only care about their own material payoff, they also care about the distribution of payoffs among players. In particular, individuals dislike having both a lower and a higher payoff than others, and so, all else equal, an equal distribution of the payoff maximizes their utility. Judging from citation figures, the inequality aversion model is very popular among experimental economists. Even though the model cannot explain the findings of some experiments, the more recent extended behavioral models that aim at explaining results in these experiments nevertheless include some concerns for equality.³

In the main part of the paper, we will test the model of inequality aversion by Fehr and Schmidt (1999, henceforth F&S). Their model has the advantage of a straightforward parametrization that can be easily estimated. Furthermore, F&S have been quite successful in rationalizing aggregate behavior in many classic games. We run four different experiments (an ultimatum game, a modified dictator game, a sequential prisoner's dilemma and a public-good game) with the same sample of experimental subjects. We then use the responder data from the ultimatum game in order to estimate a parameter of aversion to disadvantageous inequality, and we take data from the modified dictator game to estimate a parameter of aversion to advantageous inequality. Because our paper is the first to study an appropriate related-sample data set, it is also the first to report a joint distribution of individual inequality aversion parameters. We then use this distribution to test several explicit hypotheses about aggregate and individual behavior in the other games.

Our data show that results from the within-subject analysis we propose can differ markedly from results from aggregate-level tests. The inequality aversion model has considerable predictive power at the aggregate level but often fails at the individual level as

³For example, Falk and Fischbacher (2006) explicitly builds on Fehr and Schmidt (1999), while Charness and Rabin (2002) and Cox et al. (2006) assume that subjects are more altruistic towards others who have relatively low payoffs.

several of the correlations the model predicts will occur between the estimated parameters and other decisions do not materialize. That is, the degree of inequality aversion that an individual exhibits in the ultimatum game and in the modified dictator game has very little explanatory power in other games at the individual level.

The remainder of the paper is organized as follows: Section 2 presents the experimental design, followed by an instrument check in Section 3. Section 4 presents the model and the estimation of the model parameters. In Section 5 we test several hypotheses derived from the inequality aversion model. In Section 6 we discuss our findings and Section 7 concludes.

2 Experimental design

We ran four different two-person one-shot games of similar complexity with the same sample of experimental subjects. We kept the initial total surplus at £20 across all games. Each game was played exactly once by each subject. Two of the games involve two different roles for decision makers. In these games, each subject made a decision in both roles. Hence subjects made decisions in six different roles. When a role involved decisions in more than one decision node, we used the so-called strategy elicitation method to elicit choices in all these nodes.

Each of the four games was presented separately in a different section of the experiment. Instructions were distributed and were also read aloud in each of the four parts by the experimenter and participants had the chance to ask questions. Once the experimenter had ensured that everyone had understood the game, the corresponding computer screen was displayed and subjects submitted their decisions. Only when all participants had made their decisions in one game were the instructions for the following game distributed.

Subjects did not receive any feedback or payment until the end of the experimental session. All decisions were to be made without any information on other subjects' choices

and without any communication. At the end of the session, one game was chosen randomly and subjects were randomly matched in pairs. In all games (except one where symmetric players move simultaneously), the roles in the game were determined randomly between the two subjects of each pair. The payments to the subjects were determined by the single decision pair in the one game randomly chosen at the end. Subjects knew about this procedure in advance and the computer screen at the end of the experiment informed them about all the random draws of the computer and also about the payment-relevant decisions. We believe that our design is appropriate for minimizing confounding effects among games and avoiding subjects averaging their earnings across games.⁴

When selecting the games for our experiment, we wanted to make sure that we included the most relevant ones in the other-regarding preferences literature. Therefore, we chose the ultimatum game, the dictator game, and the public-good game. Finally, we used the sequential prisoner's dilemma (Clark and Sefton, 2001) because it shares crucial qualitative properties with the gift-exchange game (Fehr et al., 1993) and the trust or investment game (Berg et al., 1995) but is much simpler. We also had to decide on the number of games to be played. Four games seemed to us a reasonable compromise between generating a rich data set and maintaining salient incentives (Smith, 2002). With a higher number of games, we might have risked subjects not caring any longer about each individual decision.

We now introduce the four games as implemented in our experiment. See Table 1 for a summary of our design. The ultimatum game (henceforth UG) (Güth et al., 1982) is a sequential two-stage game. Given a pie of £20, the proposer has to make an offer (£ s) to the responder, keeping £20– £ s to himself. The responder can accept or reject the offer. In the case of a rejection both players earn zero. If the responder accepts, players get the outcome proposed, £20– £ s and £ s , respectively. As mentioned above, we let

⁴Regarding feedback and payments, our design is very similar to Charness and Rabin (2002). See their paper for a further discussion of issues arising due to the related-sample design.

subjects decide as both proposers and responders. Since we wanted to avoid feedback, the responder decisions can only be made based on a menu of hypothetical offers (this is the aforementioned strategy-elicitation method). That is, when deciding as the responder, subjects had to accept or reject a complete list of every possible distribution of the pie, starting from £20-£0, £19-£1, £18-£2, ... all the way to £0-£20. As proposers' offers were restricted to integers, there were 21 different distributions to decide upon. If the ultimatum game was selected as the game relevant for the final payment to subjects, the proposer's actual offer was compared to the responder's decision about this offer and payments were finalized according to the rules of the ultimatum game.

In the standard dictator game (Forsythe et al., 1994), the dictator unilaterally determines how to divide a fixed amount of money (£20 in our case) between himself and the recipient. The distribution chosen by the dictator is final. The standard dictator game is not suitable for getting a point prediction of the parameter measuring aversion to advantageous inequality (see F&S). Therefore, we implemented a modified dictator game (henceforth MDG) resembling more the dictator game in Kahneman et al. (1986) where dictators could only choose between allocations of (10, 10) and (18, 2). In our modification, the dictator has to decide about how much of the initial pie of £20 (if any) he is at most willing to sacrifice in order to achieve an equal distribution of payoffs. More specifically, subjects were given a list of 21 pairs of payoff vectors, and they had to choose one of the two payoff vectors in all 21 cases. The left payoff vector was always (£20, £0), that is, if the left column was chosen, the dictator would receive £20 and the recipient nothing. The right payoff vector contained equal payoffs varying from (£0, £0), (£1, £1) all the way to (£20, £20).⁵ The MDG involves a decision maker and a passive player. Each subject

⁵In this modified dictator game, a purely self-centered individual would always choose (£20, £0) over all equal payoff vectors up to (£19, £19), and would be indifferent between (£20, £0) and (£20, £20). A dictator who strongly dislikes advantageous inequality would always choose the right column with equal

made a choice in the role of the decision maker. If the MDG was randomly selected at the end of the experiment, one of the 21 payoff vector pairs was randomly chosen and then the dictator's decision determined the payments.

The sequential prisoner's dilemma (henceforth SPD) (Clark and Sefton, 2001) is a prisoner's dilemma where one player moves first, the other player second. The first mover can cooperate or defect. After observing this action, the second mover responds either with cooperation or defection. If both defect, both players receive a payoff of £10. If both cooperate, they get £14 each. If one defects and the other cooperates, players earn £17 and £7, respectively. As in the ultimatum game, subjects had to play both roles. They had to make two second-mover decisions, one if the first mover decides to defect and one if he cooperates. When the SPD was selected as the game relevant for the final payment to subjects, one subject was randomly allocated the role of first mover and the other the role of second mover. Their payoffs were then determined based on their decisions.

Finally, the public-good game (henceforth PG) we used was a simple two-player voluntary contribution mechanism (see Ledyard, 1995, for a survey). The two players received an endowment of £10 each. They simultaneously decide how much (if any) money from the endowment to contribute to a public good. Each monetary unit that the individual keeps for himself raises his payoff by exactly that amount. Both subjects receive £0.7 for each £1 contributed to the public good (this is the marginal return per capita). Note that, when restricting actions to the extreme choices of zero and full contribution, the set of possible payoffs is the same as in the SPD. If the public-good game was chosen for the final payoffs, payoffs were calculated according to the contributions of the randomly paired players.

payoffs. Subjects with monotone preferences between these two extremes should switch at some point (if at all) from choosing the left column to choosing the right column and should not switch back. The reason is that the egalitarian outcome is "cheaper" for all decisions beyond the switching point. If a player prefers (£7, £7) to (£20, £0), this player should also prefer (£8, £8) to (£20, £0) and so on.

We implemented two different sequences in which the games were played. Because of the similarity of the games, we wanted to avoid either UG and MDG or PG and SPD being played back-to-back. Also, because of the length of the instructions, we wanted the PG to be the last game. This leaves two possible sequencing variants with either the ultimatum game coming first and the dictator game coming third, or vice versa. The sequential prisoner's dilemma would be played as the second game, and the public-good game would be last. This is only a small subset of the 24 possible sequencing variants, but to run sufficiently many repetitions of all variants does not appear to be feasible.⁶ We did not find any significant differences between the two sequences and therefore we pool the data and refrain from further references to the sequences in the results section. Moreover, our results do not differ much from previous experiments where the four games were played in isolation (see the next section) which also suggests the absence of sequencing effects.

We ran six sessions with 8 to 14 subjects in each session. All 72 subjects were non-economists. In the data analysis below, we discarded 11 of those 72 subjects from the data set. The reason is that these subjects do not have a unique switching point in the MDG or no unique rejection threshold in the UG. Therefore, we cannot calculate their inequality aversion parameters and decided to drop them from the analysis. Henceforth, we will deal with a total of 61 subjects.⁷ The experimental software was developed in z-Tree (Fischbacher, 1999). Sessions lasted about 50 minutes and the average earnings were £11.

⁶Given that the ultimatum game and the SPD can be played in two and six sequences, respectively, we would even have to take 288 different variants into account.

⁷Holt and Laury (2002) elicit risk preferences with sets of binary choices similar to our UG responder decisions and our MDG. In their data, 19.8% of the subjects had a non-unique switching point, slightly more than the 15.3% we observed.

3 Instrument check

In this section, we check whether the games we analyze below generate results similar to those of previous experiments. Such an instrument check (Andreoni et al., 2003) is essential for the significance of the main part of our analysis.

In our UG, proposers' mean offer is 40% of the pie. Roughly half of the proposers (48%) offer the equal split which is also the modal and median offer. About 11% of the offers are consistent with subgame perfect equilibrium (which is to either offer nothing or £1). These results are remarkably similar to the results obtained under the standard UG design as reported in the meta study of Oosterbeek et al. (2004). They also found a mean offer of 40%, and that 50% offer the equal split. See also Roth (1995) and Camerer (2003). Regarding responder decisions, our results are consistent with the categorization in F&S (we elaborate on this extensively in the next section) which is derived from data in Roth (1995).

In the MDG, the average switching point was roughly (£11, £11). The modal switching point was (£10, £10) (with a frequency of 13%) and 43% of the subjects switched to the egalitarian outcome in the range of (£0, £0) to (£9, £9). There are 8% of the subjects who switch to the egalitarian outcome only when it is costless, at (£20, £20), and a further 10% who do not switch at all, that is, they even choose (£20, £0) over (£20, £20). Two of 61 subjects choose (£0, £0) over (£20, £0). Because we use a novel modification of the dictator game, the results cannot be directly compared with those reported for standard dictator game experiments. One parallel that can be drawn is that Forsythe et al. (1994) found that 20% of the dictators chose not to pass anything to the other player, a figure which is in line with the number of subjects in our experiment who never choose the egalitarian outcome or do so only when it is costless. Further, in Kahneman et al. (1986), 76% of dictators prefer (10,10) over (18,2) which compares with the 62% of dictators in our experiment who

switch to the equal distribution at (£12,£12) or below. These dictators pay at least 8 out of an initial pie of 20 to achieve an equal distribution like in Kahneman et al. and thus our data are roughly in line with theirs, despite the differences in procedures.

In the SPD, 34% of the subjects cooperated as the first mover. In the role of second mover, 38% cooperate following first mover's cooperation. Given first-mover defection, nearly all subjects (94%) also defected. Our results are remarkably similar to those obtained by Clark and Sefton (2001) in their SPD. The figures they obtained ("baseline" treatment, last round)⁸ are 32.5% cooperation of first movers, 38.5% second mover cooperation given first mover cooperation, and 96% defection given first mover defection.

In our PG, the average contribution was 47% of the endowment. Less than half the endowment was contributed by 41% of the subjects, including 28% (of the total population) who contributed nothing. Not contributing was also the modal behavior. More than half the endowment was contributed by 44% of the subjects, including 18% (of the total population) who contributed the entire endowment. Goeree et al. (2002) report on one-shot public-good games. They have one treatment with two players where the marginal per capita return is similar to ours (0.8).⁹ The average contribution in that treatment is 50%, very similar to our average. Roughly 47% gave less than half the endowment and 53% gave more than half the endowment. Considering that the equal split was not possible in Goeree et al. (2002), since the endowment was 25 tokens, again, the results are remarkably similar to those we

⁸Clark and Sefton (2001) repeat their SPD and report cooperation rates in the first and the last rounds. We consider the last round of their data more relevant for comparison to our one-shot setting. Moreover, the percentage gain from exploiting compared to reciprocating cooperation is 21% in our game which compares with the 20% gain in the "baseline" treatment of Clark and Sefton (2001).

⁹Most of the treatments in Goeree et al. (2002) distinguish between an internal and an external return factor. We refer to the treatment ("N=2, \$0.04, \$0.04") where both factors are equal as in standard PG experiments like ours.

observed in the PG. Differences from our results are that they observe fewer cases of zero contributions (10%) but also fewer full contributions (6%).

We conclude that our results successfully replicate those of other experiments (even in the subgames of the UG and the SPD) despite our related-sample design. Therefore, our design should be suitable for the individual-level test of the inequality aversion model.

4 Model and estimation of the parameters

In F&S' outcome-based theory, other-regarding preferences are modeled as inequality aversion. This means that players are concerned not only about their own material payoff but also about the difference between their own payoff and other players' payoffs. For two-player games, a F&S utility function is given by

$$U_i(x_i, x_j) = \begin{cases} x_i - \alpha_i(x_j - x_i), & \text{if } x_i \leq x_j \\ x_i - \beta_i(x_i - x_j), & \text{if } x_i > x_j \end{cases} \quad (1)$$

where x_i and x_j , $i \neq j$, denote the monetary payoffs to players i and j .

F&S make the following a priori assumptions on the distributions of the parameters. First, they assume $\beta_i \leq \alpha_i$, meaning that individuals suffer more from disadvantageous inequality than from advantageous inequality. Second, they impose $0 \leq \beta_i < 1$, where $0 \leq \beta_i$ rules out individuals who enjoy being better off than others and $\beta_i < 1$ excludes individuals who will burn money in order to reduce advantageous inequality. In order to rationalize the results of other experiments, F&S further assume that $\beta_i < n/(n-1)$ for $n = 6$, hence $\beta_i < 0.8\bar{3}$ (p. 832), and that α and β are positively correlated (p. 864). Finally, F&S derive a distribution of α and β (see below) which they argue is consistent with previous experimental evidence.¹⁰

¹⁰See also Shaked (2005) and Fehr and Schmidt (2005).

We follow F&S in deriving the distribution of the parameter for aversion to disadvantageous inequality, α , from the UG responder decisions. Since we employ the strategy elicitation method, the rejection thresholds in the ultimatum game give us (near) point estimates of α_i for each individual. To see this, suppose s'_i is the lowest offer responder i is willing to accept, and, consequently, $s'_i - 1$ is the highest offer i rejected (recall that choices had to be integers). It follows that this responder (assuming well-behaved preferences) is indifferent between accepting some offer $s_i \in [s'_i - 1, s'_i]$ and getting a zero payoff from a rejection. Therefore, we have $U_i(s_i, 20 - s_i) = s_i - \alpha_i(20 - s_i - s_i) = 0$. (Note that only the range of offers up to half of the pie is relevant here.)

Thus, the estimate of the parameter of disadvantageous inequality is

$$\alpha_i = \frac{s_i}{2(10 - s_i)}. \quad (2)$$

For our estimation, we set $s_i = s'_i - 0.5$. This is somewhat arbitrary but it in no way affects our results because we use non-parametric tests which are based on ordinal rankings of outcomes. A rational F&S player will always accept the equal split in the UG and hence have $s'_i \leq 10$, so division by zero cannot occur by assumption here. For a subject with $s'_i = 0$, we observe no rejected offer and we cannot infer the indifference point s_i . Therefore, we set $\alpha_i = 0$ for participants with $s'_i = 0$ but it could actually be that these subjects have $\alpha_i < 0$, that is, they could positively value the payoff of another player who is better off.¹¹ For subjects who accept only $s_i \geq 10$, we can only infer that $\alpha_i \geq 4.5$. We assign $\alpha_i = 4.5$ to these subjects. This is somewhat arbitrary but not relevant in our analysis below.

Let us now turn to the parameter of aversion to advantageous inequality, β . F&S derive the distribution of this parameter from offers in the UG. In our view, there are various problems with this. First, proposers' offers depend on their beliefs about the other players'

¹¹See Charness and Rabin (2002) and Engelmann and Strobel (2004) for evidence that at least in non-strategic games such preferences may occur.

minimum acceptance threshold in the UG. F&S assume that proposers know the empirical distribution of α . While this is a plausible way of proceeding, their conclusions on the β distribution hinge on the assumption that the beliefs are correct. Second, F&S derive the β distribution assuming risk neutrality—which may not hold for all proposers. Risk averse proposers may propose the equal split even if they do not care about inequality. Third, even a relatively small number of responders with high rejection thresholds can imply that the optimal decision of a selfish proposer ($\beta = 0$) is to offer half the endowment (this is the case in our data, see below). In that case, no β distribution can be derived because all proposers should make the same offer. Fourth and most importantly, with the method F&S use, it is only possible to derive three relatively coarse intervals of the β parameter (see below).

We prefer to derive (nearly) exact point estimates for β_i analogously to the way the α_i were derived.¹² In the UG, α_i is defined by the offer that makes responder i indifferent between accepting and rejecting the offer. In our modified dictator game, we can get a point estimate for β_i by finding the egalitarian allocation, (x_i, x_i) , such that the dictator is indifferent between keeping the entire endowment, the $(20, 0)$ outcome, and (x_i, x_i) . In the Appendix, we show that the design of our MDG is structurally the simplest design which provides a point estimate for the whole range of relevant β . Note that both the α and the β parameters are derived from non-strategic choices. Neither the UG responder nor the MDG choice depends on subjects' beliefs about how the other player is expected to play.

Suppose an individual switches to the egalitarian outcome at a payoff vector (x'_i, x'_i) . That is, he prefers $(20, 0)$ over $(x'_i - 1, x'_i - 1)$ but (x'_i, x'_i) over $(20, 0)$. We conclude that he is indifferent between the $(20, 0)$ distribution and the $(\tilde{x}_i, \tilde{x}_i)$ egalitarian distribution where

¹²As will become clear below, all decisions have implications for the model parameters. However, we refer to the parameters derived here as the α_i and β_i because they are (near) point estimates. See also Section 6.

$\tilde{x}_i \in [x'_i - 1, x'_i]$ and $x'_i \in \{1, \dots, 20\}$. From (1) we get $U_i(20, 0) = U_i(\tilde{x}_i, \tilde{x}_i)$ if, and only if, $20 - 20\beta_i = \tilde{x}_i$. This yields

$$\beta_i = 1 - \frac{\tilde{x}_i}{20}. \quad (3)$$

For our data analysis, we use $\tilde{x}_i = x'_i - 0.5$ (which, as above, does not affect the results of non-parametric tests). However this does not work well at the boundaries. Subjects who choose $(0, 0)$ over $(20, 0)$ are possibly willing to sacrifice more than £1 in order to reduce the inequality by £1. Therefore, these subjects might have $\beta_i > 1$. Since we do not observe a switching point for these subjects, we cautiously assign $\beta_i = 1$ to them in the data. Similarly, subjects who prefer $(20, 0)$ over $(20, 20)$ are possibly willing to spend money in order to increase inequality. These subjects might have $\beta_i < 0$ but, again, we do not observe a switching point for them and therefore we set $\beta_i = 0$ for such subjects in our data.¹³

In Table 2, we summarize the distributions of the α and the β parameters. The table lists both the distribution as assumed in F&S and our results. For both parameters, F&S assume few points in the density with mass (p. 844). The α density is assumed to have mass at $\alpha = 0$ (30%), $\alpha = 0.5$ (30%), $\alpha = 1$ (30%), and $\alpha = 4$ (10%). The β density function in F&S has mass at three points, $\beta = 0$ (30%), $\beta = 0.25$ (30%), and $\beta = 0.6$ (40%). For the comparison in Table 2, we prefer to interpret these mass points not literally but instead refer to the broader intervals which F&S used in their derivation (see pp. 843-4). The intervals in Table 2 for the α parameter correspond to those intervals F&S suggest for the rejection thresholds in the UG. Starting from the top segment, F&S propose the following intervals. Subjects who reject even offers that are close to an equal split; subjects who insist on getting at least one third of the pie; subjects who insist on getting at least a quarter

¹³F&S (p. 824) acknowledge that subjects with $\beta_i < 0$ may exist and indeed behavior consistent with the existence of such preferences has been observed in the experiments of Huck et al. (2001).

of the pie; and subjects who are willing to accept less than that. It is readily verified that these rejection thresholds imply the intervals for the α parameter in the table.¹⁴ Given these intervals, the α distribution we derive does not differ significantly from the one assumed in F&S ($\chi^2 = 1.79$, *d.f.* = 3, $p = 0.618$, one-sample test). At the extreme ends of the distribution, we find nine subjects with $\alpha_i = 0$ and eight subjects with $\alpha_i \geq 4.5$.

As for the β distribution, we use the very intervals F&S (p. 844) derive. (F&S only assign the aforementioned mass points within the intervals at a later stage.) The distribution of β in our data differs significantly from the one in F&S ($\chi^2 = 8.51$, *d.f.* = 2, $p = 0.014$, one-sample test). We find seven subjects (11%) with $\beta > 0.8\bar{3}$, two of which have $\beta_i = 1$. We also observe six subjects with $\beta_i = 0$.

A key novelty of our data set is that we can estimate the *joint* distribution of α and β . Previous research, including F&S, could not derive the joint distribution because related-sample data were not collected. Figure 1 shows this joint distribution. Both parameters turn out to be widely distributed in the population. It is apparent that the α_i and β_i are not significantly correlated and the Spearman correlation coefficient confirms this ($\rho = -0.03$, $p = 0.820$). We find that 23 of our 61 subjects violate the F&S assumption that $\alpha_i \geq \beta_i$. They can be found to the left of the $\alpha = \beta$ line in the figure.

To summarize, our method for deriving the α and β distribution has, to a large extent, replicated the distribution chosen in F&S. Even though our β distribution is significantly different from the one in F&S, the distributions do not differ grotesquely, and indeed our

¹⁴The reader will note that the top interval in Table 2 corresponds to $4.5 \leq \alpha < \infty$ whereas F&S assign $\alpha = 4$ in the segment of subjects with the highest α . The reason for this discrepancy is that F&S' description of this interval (subjects who reject "offers even if they are very close to an equal split") applies best to those responders in our data who accept only £10 or more. Since these subjects reject an offer of £9, they get $\alpha_i \geq 9/(2(10 - 9)) = 4.5$. This value is only slightly higher than the $\alpha = 4$ F&S assign and, moreover, F&S consider their own estimate "conservative".

β distribution is rich enough to conduct meaningful tests of the model. In a way, this can be seen as support of the distributions F&S assume at the aggregate level. However, the joint distribution of α and β does not support two of the assumptions F&S make at an individual level ($\alpha_i \geq \beta_i$, positive correlation of α_i and β_i).¹⁵

5 Tests of the inequality aversion model

We now move on to test several hypotheses derived from the F&S model. Formal derivations of the hypotheses are presented in the Appendix. We will analyze the results for a game in two steps. We will first assess the predictive power at the aggregate level and second at the individual level. The aggregate-level analysis will ignore the within-subject character of our data and the analysis will be *as if* the data on the inequality aversion parameters and those on the other decisions came from unrelated experiments. This is how previous tests of the F&S model have proceeded. We will then go beyond that approach by analyzing the individual-level data.

This is how we will conduct our tests formally. Like with the example in the introduction, consider the fictitious hypothesis “subject i will choose the altruistic action in some game G if, and only, if $\alpha_i \leq \bar{\alpha}$ ”, where $\bar{\alpha}$ is some numerical threshold derived from the model. We will accept this hypothesis at the aggregate level if a chi-square test of proportions does

¹⁵Our UG design explicitly asks for acceptance or rejection of each possible offer. Interestingly, we observe seven subjects who consistently reject offers $s \geq s'$ for some $s' > 10$. Since $s > 10$ here, these decisions also reveal that these subjects have a high degree of aversion towards advantages inequality (actually, $\beta_i > 1$). For these seven subjects, we find some relation to the β_i we estimate from the MDG. The average β_i of these subjects is higher than that of the rest of the sample (although this is not significant at the five-percent level). For two of these subjects we find $\beta \geq 1$ based on the MDG. In any event, responders could expect the probability of receiving such an offer to be close to zero, so that their decisions are effectively cheap talk.

not indicate any significant difference between the share of subjects with $\alpha_i \leq \bar{\alpha}$ and the share of altruistic subjects in game G . At the individual level, the hypothesis is confirmed if a chi-square test indicates that the proportion of subjects with $\alpha_i \leq \bar{\alpha}$ who make the altruistic choice is significantly larger than the proportion of subjects with $\alpha_i > \bar{\alpha}$ who do so.

The hypotheses we derive from F&S are sometimes not unconditional as in the example but depend on the beliefs players hold about the inequality aversion (and resulting behavior) of the other players. In those cases, we will first test what should happen when subjects have correct beliefs about the distribution of inequality aversion parameters in the sample. This is also the assumption underlying the analysis in F&S. Second, where appropriate, we derive some auxiliary hypotheses at the individual level for arbitrary random beliefs which are not correlated with players' types.

As for the statistical tools for our tests, we will almost exclusively apply non-parametric tests. In addition to the chi-square test of proportions already mentioned, we will sometimes use correlation analysis (Spearman correlation for two ordinally scaled variables; rank biserial correlation for one ordinally scaled variable and one dichotomous nominally scaled variable; phi coefficient for two dichotomous nominally scaled variables). Non-parametric tests interpret the data in an ordinal fashion which we consider appropriate here. Because non-parametric tests use only ordinal rankings, they may apply equally to possible non-linear generalizations of F&S, an issue we discuss below.

5.1 Offers in the ultimatum game

Our main hypothesis for the ultimatum game is as follows.

Hypothesis 1 (i) *Subjects with $\beta_i > 0.5$ should offer $s_i = 10$ in the Ultimatum Game.* (ii) *Subjects with $\beta_i < 0.5$ may, depending on their beliefs, offer either $s_i = 10$ or $s_i < 10$ in the Ultimatum Game.*

Consider part (i). We take a look at the aggregate level first and compare predictions and data as if they came from different data sets and without taking the available within-subject information into account. In the data, we have 33 subjects with $\beta_i > 0.5$ and 26 subjects with $\beta_i < 0.5$.¹⁶ In the UG, we observe 29 subjects who offer $s = 10$. The aggregate outcome of $s = 10$ offers is not inconsistent with F&S since subjects with $\beta_i < 0.5$ should offer $z < 10$ for some beliefs. The deviation of actual from the predicted $s = 10$ observations is $(33-29)/33 = 12.1\%$ which seems small enough to consider the F&S prediction reasonably accurate. More formally, we cannot reject that the share of subjects with $\beta_i > 0.5$ is the same as the share of subjects offering $s = 10$ ($\chi^2 = 0.544$, $d.f. = 1$, $p = 0.461$). At the aggregate level, we can accept Hypothesis 1 (i).

At the individual level, the data do not support the F&S model. Among the 33 subjects with $\beta_i > 0.5$, 18 chose $s = 10$, only slightly more than half of this group. A chi-square test on the $\beta_i > 0.5$ observations cannot reject that choices are equiprobable ($\chi^2 = 0.273$, $d.f. = 1$, $p = 0.602$, one-sample test¹⁷). Robustness checks with various thresholds $\beta \in [0.3, 0.7]$ reveal that the insignificance of the result does not depend on the particular value of the $\beta = 0.5$ threshold. Therefore, we reject Hypothesis 1 (i) at the individual level.

As for the second part of the hypothesis, among the 26 subjects with $\beta_i < 0.5$, 11 chose $s = 10$. The individual behavior of these subjects is consistent with Hypothesis 1 (ii) if subjects hold heterogenous beliefs¹⁸ but it seems remarkable that the share of subjects

¹⁶There are two subjects in the sample who offer $s > 10$. These subjects are not consistent with F&S regardless of their β parameter. Therefore, we cannot interpret their UG offer within the inequality aversion model and so we discard them from the analysis. Note also that $\beta_i = 0.5$ for no subject in our sample, so, we only need to distinguish $\beta_i \gtrless 0.5$.

¹⁷We cannot apply a two-sample test here (that is, testing the $\beta_i > 0.5$ part of the sample versus the $\beta_i < 0.5$ part) because we do not have an unconditional hypothesis for the $\beta_i < 0.5$ subjects.

¹⁸Note that, if we claim that the behavior of the $\beta < 0.5$ subsample is consistent, the aggregate outcome no longer supports F&S. The reason is that, if the $s = 10$ choices of $\beta < 0.5$ subjects are to be rational F&S

offering $s = 10$ here does not differ significantly from the one observed for the $\beta_i > 0.5$ subjects ($\chi^2 = 0.871$, $d.f. = 1$, $p = 0.351$). Figure 2 graphically displays the findings on Hypothesis 1 at the aggregate and individual level.

Assume now that subjects know the true distribution of α . In that case, it turns out that all subjects should offer $s = 10$ in the Ultimatum Game, regardless of their β_i . This hypothesis is clearly rejected from what was said above. On the one hand, this might result from subjects' beliefs being wrong. On the other hand, it could be that the behavior of proposers is not accurately captured by the β parameter. Furthermore, this observation confirms our decision not to derive the β parameter from the ultimatum game. From our UG proposer data, no β distribution can be derived.¹⁹

choices, we should observe 44 $s = 10$ choices in total, which is substantially different from the 29 actual observations, with a deviation from the prediction of $(44 - 29)/44 = 34.1\%$. In other words, the degree of freedom arising due to arbitrary beliefs about responder behavior can only be used to either rationalize the outcome at the aggregate level or the behavior of the $\beta < 0.5$ subsample.

¹⁹The result may also indicate that the assumption of risk-neutral behavior underlying F&S' analysis is not appropriate here. Expected payoffs from offering $s > 5$ are relatively flat. Assuming $\beta_i = 0$ and given the responder behavior in our data, offering $s \in \{6, 7\}$ yields 9.0 on average, offering $s \in \{8, 9\}$ yields an expected payoff of roughly 9.6, and offering the equal split always yields a payoff of 10. Given these very similar expected payoffs, small differences in risk attitudes might provide a more plausible explanation for the heterogeneity of UG offers than inequality aversion does. The flat expected payoffs also highlight the point made above that UG offers are not suitable for deriving the distribution of β parameters. Suppose, in contrast to what we find, that the highest expected utility occurred (for a $\beta_i = 0$ player) for some offer $s < 10$ just as F&S assume. Even though F&S's derivation of the β parameter could, in principle, be applied, Harrison's (1989) "flat maximum" critique would apply. This critique suggests that deviations from the actual maximum should be interpreted with caution when payoffs do not differ much around the maximum.

Whereas the previous argument highlights the role of beliefs for proposer behavior, it is possible to make a prediction for general uncertain proposer beliefs. It is easy to show that UG offers of subjects with $\beta_i < 0.5$ should be positively correlated with β_i , at least as long as beliefs (concerning the rejection probability) are not systematically negatively correlated with β . The Spearman correlation coefficient is not significant ($\rho = 0.187$, $p = 0.350$)²⁰, however. (We restricted the test to the subjects with $\beta_i < 0.5$ because the other subjects should offer $s = 10$ anyway, so no correlation should occur. If we nevertheless include the subjects with a β_i larger than 0.5 in the correlation analysis, the result does not change. See Table 3 below.) We conclude that the β data have explanatory power regarding the UG offers at the aggregate level but not at the individual level.

5.2 Contributions to the public good

Hypothesis 2 (i) *Subjects with $\beta_i < 0.3$ should not contribute in the PG.* (ii) *Subjects with $\beta_i > 0.3$ may, depending on their beliefs, contribute any amount between zero and their entire endowment.*

We start with part (i) analyzing the data at the aggregate level. There are 20 subjects with $\beta_i < 0.3$ and we observe 17 subjects who contribute zero. The data at the aggregate level are consistent with F&S if we assume that all 41 subjects with $\beta > 0.3$ believe the other player will contribute as well (there are no subjects with $\beta = 0.3$). The formal test suggests that the proportion of zero contributors is not significantly different from the one of $\beta_i < 0.3$ subjects ($\chi^2 = 0.349$, $d.f. = 1$, $p = 0.55$). Following Andreoni (1995), one could argue that merely positive but small contributions in the PG result from confusion and do not indicate a true intention to cooperate. Therefore, we alternatively consider subjects who contribute less than half of the endowment as non-contributors. There are 25 subjects

²⁰We report two-tailed p values throughout.

who do contribute less than half of their endowment. Again, this is consistent with F&S at the aggregate level ($\chi^2 = 0.88$, $d.f. = 1$, $p = 0.348$).

At the individual level, among the 20 subjects with $\beta_i < 0.3$, 13 [10] choose a positive contribution [at least half their endowment]. This is not consistent with F&S. A chi-square test on the $\beta_i < 0.3$ observations cannot reject that the proportions of zero versus positive contributors are equiprobable ($\chi^2 = 1.80$, $d.f. = 1$, $p = 0.18$, one-sample test), where the deviation from equiprobable choices is opposite to the prediction. The proportion of contributors of less than half the endowment is exactly 10 out of 20 and therefore not significantly different from being equiprobable ($\chi^2 = 0.00$, $d.f. = 1$, $p = 1.00$, one-sample test). Among the subjects with $\beta \geq 0.3$, 31 of 41 made a positive contribution, and 26 contributed at least half the endowment. This outcome is consistent with F&S. However, the difference from the subjects with $\beta_i < 0.3$ is not significant when considering both merely positive contributions ($\chi^2 = 0.75$, $d.f. = 1$, $p = 0.39$) and contributions of at least half of the endowment ($\chi^2 = 1.00$, $d.f. = 1$, $p = 0.32$). As robustness checks, we analyzed various levels of contributions to the PG and various thresholds of β . None suggested a significant explanatory power of the β parameter at the individual level (for example, the share of subjects who contribute the entire endowment of 10 to the PG is almost identical for the $\beta_i \geq 0.3$ subpopulations, 3 out of 20 ($\beta_i < 0.3$) and 8 out of 41 ($\beta_i > 0.3$), respectively). Figure 3 summarizes these results.

Next, we can check whether PG contributions are more basically correlated with the α and β parameters. Straightforward reasoning (see also F&S) suggests that contributions to the PG of subjects with $\beta_i > 0.3$ should be negatively correlated with α_i and positively correlated with β_i . The intuition behind the hypothesis is that, the higher α_i , the more subject i suffers from being exploited in the PG. Hence, if a subject is uncertain about the contribution of the other subject, a higher α_i makes the subject contribute less or even nothing. The opposite holds for the advantageous inequality parameter, β_j . However, we

cannot find support for the model at the individual level here either. Spearman correlation coefficients indicate that the correlations have the right sign but neither the correlation between α_i and contributions ($\rho = -0.177, p = 0.268$) nor that between β_i and contributions ($\rho = 0.104, p = 0.520$) are significant for the $\beta_i > 0.3$ subjects.²¹ The results do not change in favor of the F&S model when we also include the subjects with $\beta_i < 0.3$, or when we include only subjects with some higher $\beta_i > \tilde{\beta} \in [0.3, 0.6]$.²²

Finally, we consider the case where subjects know the true joint α - β distribution. If this is the case, no subject should contribute to the PG.²³ From the fact that 44 subjects contribute a positive amount, we conclude that F&S does not provide an accurate joint representation of both subjects' behavior and beliefs, but we cannot distinguish whether it

²¹If we take a look at the more extreme choices in the PG data, we even find that subjects who have an $\alpha_i > 2$ are more likely to contribute the full endowment compared to the rest of the sample ($\chi^2 = 5.645, d.f. = 1, p = 0.018$) whereas the F&S model predicts the opposite.

²²Since α and β influence the optimal level of contributions simultaneously, we also ran a simple least squares regression with the level of contribution as dependent variable and both α and β as independent variables. Again the impact of both inequality parameters is far from significant ($p = 0.843$ and $p = 0.565$ for α and β , respectively). The same holds for probits for the decision to contribute either more than zero, at least half or all of the endowment.

²³The proof is by iterated elimination of dominated strategies. First, note as above that for the 20 subjects with $\beta < 0.3$ contributing nothing is strictly dominant. Knowing that, the remaining 41 subjects face a player who contributes zero with probability $\geq 1/3$. Hence, making a positive contribution is dominated for subjects with $2\beta/3 - \alpha/3 < 0.3$ since it reduces advantageous inequality with probability $\leq 2/3$ but increases disadvantageous inequality with probability $\geq 1/3$. This is true for 28 of the remaining subjects. Eliminating dominated strategies, the remaining subjects face with probability $\geq 4/5$ a player who contributes 0 and hence making a positive contribution is dominated if $\beta/5 - 4\alpha/5 < 0.3$ which is true for all remaining players. Hence, all players contributing nothing is the only (Bayesian) equilibrium.

does not capture behavior or beliefs (or both).²⁴

5.3 Second move in the SPD

By backward induction, we start analyzing the SPD with second-mover behavior. Note that the next hypothesis does not depend on beliefs. Therefore, behavior should unconditionally depend on the inequality aversion parameters only, making the analysis simpler than the above cases.

Hypothesis 3 *(i) Given first-mover cooperation, second movers in the SPD should defect if, and only if, $\beta < 0.3$. (ii) Given first-mover defection, second movers in the SPD should defect.*

Consider the aggregate level first. Regarding part (i) of the hypothesis, we have 20 subjects with $\beta_i < 0.3$ in the data but we have 38 subjects who defect given first mover cooperation. Prediction and experimental data differ by $(38 - 20)/20 = 90\%$. The hypothesis that the proportion of $\beta_i < 0.3$ players is identical to the proportion of defectors is rejected ($\chi^2 = 10.65$, $d.f. = 1$, $p = 0.001$). As for part (ii), subjects should defect given first-mover defection and indeed 57 out of 61 subjects did so. While this strongly supports F&S, we

²⁴It is also possible to test a different hypothesis, namely that subjects know the true distribution of PG contributions and play their F&S best response. We numerically derived each subject's optimal contribution given their α and β . On the aggregate level, results do not confirm F&S predictions. They predict 20 positive contributions whereas we observe 44. This difference is highly significant ($\chi^2 = 18.93$, $d.f. = 1$, $p < 0.001$). The model does have some limited predictive power even though the Spearman correlation of predicted and observed contributions is not significant at the five percent level ($\rho = 0.22$, $p = 0.093$). The result is driven by the fact that the F&S prediction only rarely predicts a positive contribution but, if it does, it is quite often right. Subjects predicted to have a positive contribution typically have a high β_i . Indeed, all but two subjects predicted to make a positive contribution violate the $\alpha_i \geq \beta_i$ assumption of F&S.

note that F&S makes the same prediction here as the standard theory of rational payoff maximization.

Interestingly, even though F&S fails to explain choices at the aggregate level in part (i), the individual β_i have predictive power regarding second mover decisions when first movers cooperate. When $\beta < 0.3$, 16 out of 20 subjects defect whereas, when $\beta > 0.3$, “only” 22 out of 41 defect. This difference in cooperation rates is significant ($\chi^2 = 3.97$, $d.f. = 1$, $p = 0.046$). As above, we look for basic correlations between decisions here, and the rank biserial correlation between the cooperation decision and β supports part (i) of the hypothesis ($r_{rb} = 0.341$, $p = 0.007$). Part (ii) of the hypothesis is strongly supported also at the individual level as virtually all subjects decided according to the F&S theory. We conclude that F&S has predictive power at the individual level but not at the aggregate level for second movers in the SPD, and we summarize this finding in Figure 4.

5.4 First move in the SPD

We finally turn to the first movers. First-mover behavior depends on the beliefs of the subjects about whether or not second movers will reciprocate cooperation. If subject i believes with probability one that the second mover will reciprocate cooperation, i should cooperate regardless of the inequality aversion parameters. Similarly, if i believes that the second mover will exploit cooperation, i should defect as well. Hence, if subjects hold degenerate beliefs, the α and β parameters do not imply a hypothesis on first mover behavior. We therefore start by assuming correct beliefs here.

Hypothesis 4 *If subjects know the true distribution of the β parameter, first-movers in the SPD should cooperate if, and only if, $\alpha_i < 0.52$.*

In the data, we have 30 subjects with $\alpha_i < 0.52$ and we have 21 subjects who cooperate as first movers. The share of $\alpha_i < 0.52$ subjects and first mover cooperators does not differ

significantly at the five-percent level ($\chi^2 = 2.729$, $d.f. = 1$, $p = 0.099$). While the F&S prediction is not perfectly accurate here, it is not completely off the mark either. However, at the individual level the data is not consistent with F&S. Among the 30 subjects with $\alpha_i < 0.52$, 10 cooperate as first movers. The share of cooperators is virtually identical for the subjects with $\alpha_i > 0.52$ (11 of out of 31) so the χ^2 test does not suggest predictive power of the model at the individual level ($\chi^2 = 0.031$, $d.f. = 1$, $p = 0.86$). See also Figure 5.²⁵

A simple test for correlations does not suggest any predictive power of the model on the individual level either. If we assume alternatively that first movers' beliefs are random, then first-mover cooperation decisions and α_i should be negatively correlated. The intuition is the same as in the PG. First movers with a higher α are more averse towards being exploited and hence require a higher probability of second mover cooperation in order to cooperate. The rank biserial correlation coefficient of individual i 's first-mover "cooperate" decision and α_i is, however, practically zero ($r_{rb} = -0.032$, $p = 0.806$). It appears that aversion against disadvantageous inequality does not have explanatory power regarding first-mover behavior at the individual level even though it predicts the aggregate level reasonably well.

5.5 Correlations across games

We conclude this section by reporting correlations across all decisions of the experiment. This is done, first, for the sake of completeness and, second, because we want to exclude the possibility that individual behavior shows no systematic patterns at all across games. A reason for this could be that participants are confused by the multi-game setting and just

²⁵Alternatively, we could assume that subjects know the true distribution of second mover choices instead of the true distribution of the β parameter. In that case, no first mover should cooperate (see the proof of Hypothesis 4 and note that we have 23 subjects who cooperate as second mover; this implies that first movers cooperate only if $\alpha_i < -0.06$ which cannot hold for any subject). This hypothesis is rejected since 21 subjects cooperate as first movers.

play random choices. Or they might feel an irrational need to vary their choices, behaving fairly or cooperatively in one game and then behaving selfishly in the next. In any event, if individual behavior turned out to be completely random across decisions, the inequality aversion model could hardly be blamed for failing to predict individual decisions well.

Our data does exhibit clear patterns. Table 3 presents the correlation coefficients across the decisions made in the experiment. (We exclude the second move in the SPD given first-mover defection in Table 3 because virtually all subjects defect in this case and, hence, this decision cannot reveal any insightful correlations). Each cell contains the appropriate correlation coefficient and significant correlations are indicated with asterisks. We observe five significant correlations (plus one more if we consider PG distributions as a dichotomous variable, see below) which allows us to conclude that behavior is not random or irrationally varied across decisions.

Are the observed correlations intuitive and are they consistent with F&S? One of the five correlations (second move in the SPD and β) we have noted and discussed above already. Three of the four remaining correlations concern the second move in the SPD (given first-mover cooperation). This decision is positively correlated with UG offers, the first move in the SPD and contributions to the public good. Players should cooperate as second movers (given the first mover cooperates) if, and only if, $\beta_i > 0.3$, that is, the second move in the SPD is a good indicator of aversion to advantageous inequality. This implies that the correlations of second mover decisions with UG offers and PG contributions are consistent with F&S as both are associated with a high β . The correlation of first and second mover decisions in the SPD is difficult to reconcile with the inequality aversion model.²⁶ This

²⁶A F&S player will cooperate at both stages in the SPD if, and only if, $\tilde{\alpha} > \alpha_i \geq \beta_i > 0.3$ (for the derivation of $\tilde{\alpha}$, see the proof of Hypothesis 4). A F&S player will defect at both stages in the SPD if, and only if, $\alpha_i \geq \tilde{\alpha}$ and $\beta_i \leq 0.3$. Hence, if we elicited the α_i and β_i from these decisions, this would imply a negative correlation of the inequality aversion parameters which is inconsistent with the F&S model. Recall

correlation is, however, consistent with a consensus effect²⁷ which implies that cooperating second movers expect higher cooperation rates.²⁸

Another correlation we find is between UG offers and α_i . This result is to be expected if we assume that subjects' beliefs show a consensus effect which implies that, all other things equal, a proposer with a higher α will expect generally higher rejection rates which will (weakly) increase his utility-maximizing offer. This hypothesis is clearly supported by the strong correlation in Table 3. Andreoni et al. (2003) made the same comparison and found a similar correlation (see their “standard” treatment). The correlation between UG offers and α_i is consistent with F&S but it does not confirm any prediction of the model either. We conclude that, in addition to differences in risk attitudes, differences in expectations about the behavior of the responders can explain the variation in UG offers. Finally, the correlation between UG offers and α_i suggests that, if it was feasible to derive the β parameter from UG offers (which is not the case in our data), then the α and β parameters would be positively correlated as F&S assume.

The correlation between PG contributions and the SPD first-mover decision misses the five-percent significance level ($r_{rb} = 0.244$, $p = 0.058$) but a chi-square test with a dichotomous decision in the PG (contribute zero or a positive amount) and SPD first-mover decision is significant ($\chi^2 = 5.36$, $d.f. = 1$, $p = 0.011$). In the F&S model, both the first mover and second mover decisions are significantly correlated with the PG decision, that 45 out of 61 subjects made the same decisions as first and second movers.

²⁷In the social psychology literature the so-called “false consensus effect” is well-established (see Mullen et al., 1985). Since the label “false” is misleading because such beliefs are in principle consistent with Bayesian updating (see Dawes, 1989) “consensus effect” is a more appropriate term. See Engelmann and Strobel (2000) for evidence that subjects in an experiment with monetary incentives exhibit a clear consensus effect but no truly false consensus effect.

²⁸If we assume that beliefs are subject to a consensus effect, then first mover cooperation should be positively correlated with β , because a higher β implies a higher expectation of second mover cooperation. As seen in Table 3, this correlation is virtually zero.

move in the SPD and PG contributions are associated with a low α parameter. Therefore this correlation confirms F&S—although we note that neither decision is correlated with the α parameter.

As for consistency of choices across games, it is also instructive to look at single individuals. About one third of our subjects (20 out of 61) behave consistently in the UG, SPD and PG in the following sense. There are nine subjects who offer the equal split in the UG, cooperate both as first and as second mover in the SPD, and contribute at least half of their endowment in the PG. We also observe eleven participants offering less than the equal split in the UG, defecting at both moves in the SPD and contributing less than half of their endowment on the PG.²⁹ We think it conceivable that these subjects perceive some behavioral norm and that they conform to the norm or violate it but, whatever explains the choices of these subjects, some of their behaviors are in line with inequality aversion, others are in contrast to it. Offering the equal split [less than the equal split] in the UG and cooperating [defecting] as the second mover is associated with a high [low] β . This is confirmed in the data (even though there are no significant differences according to median tests). Cooperating [defecting] as the first mover and [not] contributing to the PG is associated with a low [high] α . This is not supported by the data as the group of subjects defecting in the cooperation games and offering less than the equal split in the UG have a lower α_i (median test, $\chi^2 = 4.234$, $d.f. = 1$, $p = 0.040$) compared to the rest of the sample, and the subjects in other group have (insignificantly) higher α_i . The point here is not to reiterate the lack of explanatory power of the α parameter in these games but to note, first of all, that a sizable share of our sample appears to exhibit stable preferences and, second,

²⁹Consistent here does not mean consistent with F&S, but with a plausible behavioral norm. A rational F&S player may well offer the equal split in the UG and cooperate as the second mover but defect as the first mover and in the PG (and vice versa). However, there are only two and one subject(s), respectively, who behave like that.

that even the behavior of subjects who make consistent choices across games is not always captured by inequality aversion. The behavioral norm underlying this behavior could be to be generally cooperative but to also demand this from others, that is, to have a high rejection threshold in the UG. This behavior partly indicates a strong degree of inequality aversion, partly just the opposite.

6 Discussion

In this section, we discuss possible explanations for our findings. We found that F&S is by and large consistent with the data from UG proposals, PG contributions and the first move in the SPD at the aggregate level but not at the individual level. For second-mover behavior (given first-mover cooperation) in the SPD, the model had predictive power at the individual level but not at the aggregate level. How can we account for these results?

Our general view of these results is that the success of the inequality aversion model at an aggregate level may be based on its ability to qualitatively capture different relevant motives in different games but that the low predictive power of the model at an individual level is driven by the low correlation of these motives within subjects. Apparently, while subjects might follow some idea of fairness across various games, this is not systematically captured by the inequality aversion parameters at the individual level.

At this point, it is important to note that F&S do not only regard distributional concerns per se as important driving behavioral forces, but also intentions or reciprocity. The α and β parameters “can be interpreted as a direct concern for equality as well as a reduced-form concern for intentions” (F&S, p. 853). The reason why inequality aversion may well capture reciprocity is that, in most experiments, both motives coincide (consider the ultimatum game where rejecting a low offer reduces the inequality of payoffs and corresponds to negative reciprocity). Closely related to the emphasis on reciprocity, Fehr and Schmidt

(2006) suggest that other-regarding preferences should be derived from a “strategic situation”. They define a strategic situation as one where the recipient of a gift can affect the material payoff of the sender of the gift. Obviously, reciprocity can only play a role in a strategic situation.

Indeed, the distinction between strategic and non-strategic situations helps considerably in understanding our results. Our α parameter is derived from a strategic situation whereas the β parameter is not. We found that the β derived from the MDG is correlated neither with PG contributions nor UG offers but that the second move in the SPD (given first-mover cooperation) is positively correlated with these decisions. Whereas the β_i do measure literal inequality aversion, they cannot capture reciprocity. By contrast, the second move in the SPD is certainly a strategic situation and also one where reciprocity matters. Hence, it appears that PG contributions and UG offers are not so much driven by literal inequality aversion³⁰ but by reciprocity and expectations of reciprocity, and—if we see their model as a shortcut for reciprocity—this is fully consistent with F&S.

However, the interpretation of the F&S model as a reduced form for reciprocity does not explain all our findings. The α parameter does not have predictive power in the PG and in the first move of the SPD even though the α_i are derived from a strategic situation where reciprocity plays a role. Also, both the first and the second moves in the SPD are presumably driven by reciprocity and expectations of reciprocity, and the two decisions are positively correlated but, as noted above, the inequality aversion parameters cannot really explain this well. Further, if we regard F&S mainly as a model of reciprocity, this raises the question of when “pure” distributional motives will still play a role. For example, behavior in the second move in the SPD is predicted well by the β parameter. While this is consistent with F&S, the systematic distinction between strategic and non-strategic

³⁰It has been suggested before that UG offers are not driven by inequality aversion or altruism but that players behave strategically (Forsythe et al., 1994; Camerer, 2003, p. 56).

situations is blurred here. It appears that both reciprocity and distributional moves are important in the SPD. Put differently, to fully account for our data, one would need to allow for subjects having both a “distribution” (or non-strategic) β and a “reciprocity” (or strategic) β , where only the former matters in MDG, only the latter in UG and PG, but both for SPD second move.

Camerer (2003, p. 56) proposes a distinction that is similar to the one between strategic and non-strategic situations suggested by Fehr and Schmidt (2006). He writes

“I suspect that Proposers behave strategically in ultimatum games because they expect Responders to stick up for themselves, whereas they behave more fairly-mindedly in dictator games because Recipients cannot stick up for themselves. This behavior could be codified in a theory of reciprocal fairness that includes responsibility.”

Camerer goes on to define the last-moving player who affects some player i 's payoff as the one “responsible” for i . If that responsible player is not player i then this player must take some care to treat i fairly. Otherwise, the player can treat i neutrally and expect i to be responsible for himself. According to this definition, both our α_i and β_i are derived from decisions where the decision maker is responsible. Like F&S' concept of a strategic situation, this notion of responsibility can explain the fact that PG contributions and UG offers are not correlated with β_i . Additionally, it can explain why neither the PG contributions nor the first move of the SPD are correlated with the α_i as subjects are responsible when deciding about the UG offers but not in PG and SPD (first move). Finally, players are responsible both in the second move in the SPD and in the MDG, and this is consistent with the positive correlation between the two choices. Here, the question arises why the second move in the SPD is also correlated with PG and UG offers where the player is not responsible.

Characterizations of decisions according to whether they are strategic or according to whether subjects are responsible illustrate that different situations might trigger different behavior. At an aggregate level, a theory based on just one motive might still be relevant—in particular if, as it seems to be the case with inequality aversion, it can account for various behavioral forces in an “as if” manner. At the individual level, in spite of the multiplicity of motives, we can still confirm a model if the same motives are relevant (for example, the second move of the SPD is consistent with PG, UG offers and β) but contradictions can easily arise (for example, the α parameter does not predict the PG and the first move of the SPD). We would hence expect a model calibrated on decisions in one type of game to yield reliable predictions only within the class of games where the same motives dominate. Since this is difficult to know ex-ante, deriving predictions for new games appears to be problematic.³¹

Concluding, our findings suggest that, in addition to the heterogeneity of subjects along one dimension (say, inequality aversion), the multiplicity of behavioral motives gives rise to a multi-dimensional heterogeneity that is difficult to account for in a simple model. For example, surely not *all* subjects ignore distributional motives when making a strategic choice. However, it is also clear that for some subjects inequality aversion is dominated by other concerns when making a strategic choice. As a result, for UG proposals, differences in expectations or risk aversion appear to dominate differences in concerns for equality.

Throughout our analysis, we have assumed that preferences are stable and that subjects are rational. Our test is, hence, one of a joint hypothesis that the model correctly describes subjects’ preferences, that these preferences are stable and that subjects are rational. Of course, it could be the case that subjects’ preferences are not stable or that participants are

³¹For example, as discussed above, a β parameter estimated based on SPD second mover behavior would have predictive power for behavior in MDG, UG and PG. Why it works better than, say, estimates derived from UG offers is, at least ex ante, not clear.

not fully rational. Regarding the instability of preferences, indeed, the social psychology literature generally reports low predictability of how an individual will behave in a given situation from past behavior and concludes that the specifics of the situation are important for individual decisions (Ross and Nisbett, 1991). A potential lack of stability of subjects' preferences might play a role in our results. Our data does, however, exhibit quite some consistency and, in some cases, substantial correlations between decisions. This may be due to the fact that subjects made the choices within a relatively short time period and under identical conditions. It appears to us that the major factor underlying the partial failure of the model at the individual level is the multiplicity of relevant behavioral forces rather than their instability.

This distinction may be partly a matter of taste. While one could, for example, argue that altruism is not stable if a subject behaves altruistically in one game but not in another, we suggest that a more appropriate explanation is that there is not such a simple motivation as altruism, but that there are several very specific motives (such as altruism towards those who are particularly poor, towards those who have been kind to me or towards those who have been kind to others) and that these are not necessarily correlated. This suggests that there will be higher stability of behavior within similar groups of games where the same specialized motive is more likely to matter. The result of Andreoni and Miller (2002) that almost all subjects make choices consistent with some convex preferences across several dictator games points in this direction.³²

As for subjects' rationality, several commentators on our paper encouraged us to check whether Quantal Response Equilibrium (McKelvey and Palfrey, 1998) can explain our data.

³²This is not to say that there are not other sources for instability of preferences, such that subjects make different choices even in the same game if the situation has changed in other respects. Having played all games in the same setting in one go as in Andreoni and Miller (2002), however, is likely to have minimized such effects in our experiment.

While we are sympathetic to this idea, we believe it is virtually impossible to conduct such an analysis here. Almost all subjects have different preferences (in terms of the α and β parameters). The QRE approach would require here that a player i chooses his (noisy) best response against a probability distribution over the types of players i faces. Now, for each type of player in this distribution, there is another probability distribution across the various choices that this type of player might take. Conducting a QRE analysis, therefore, seems a formidable task.

We note that we have applied the linear inequality model as suggested in F&S. One might suspect that a generalized non-linear version would perform better in the analysis of individual behavior. However, our main conclusions are based on the absence of a correlation between the inequality aversion parameters which we estimated based on behavior in the UG and MDG and the behavior in the other decision nodes. In a generalized version a lower (higher) switching point in MDG (UG) would still result in a stronger measure of inequality aversion. Put differently, according to the non-parametric measures we use, the switching point and the measure of inequality aversion are perfectly correlated. Furthermore, inequality aversion has again the same implications for the decisions in the other games, for example, stronger aversion to disadvantageous inequality implies lower contribution levels in the PG. Now, the absence of a correlation between our estimates of the inequality aversion parameters in the linear model and other decisions only means that the switching points in the MDG and UG are not correlated with the other behavior. Since the switching points are perfectly correlated with any non-linear measure of inequality aversion, this means that such measures would not be significantly correlated with the behavior in the other decision nodes either. Hence a non-linear model would fail to find support exactly in the same instances as the linear model and nothing would be gained from such a generalization within our framework—which is not to say that generalization would not

improve the predictive power of F&S in other instances.³³

7 Conclusions

In this paper we assess the predictive power of one of the central models of the other-regarding preferences literature—inequality aversion—using a within-subjects design. We run four different experiments (an ultimatum game, a modified dictator game, a sequential prisoner’s dilemma and a public-good game) with the same sample of experimental subjects. This allows us to make within-subjects comparisons across the decisions in the experiments. We use the responder data from the ultimatum game and the data from the modified dictator game to estimate the two parameters of the Fehr and Schmidt (1999) model. We then use this joint distribution of parameters to test several hypotheses about aggregate and individual behavior in the other games.

The data show that results from a within-subject analysis can differ markedly from results obtained from an aggregate level analysis. We found that the Fehr and Schmidt (1999) model predicts the ultimatum game proposals well at the aggregate level but not at the individual level. The same holds for contributions to the public good and the first move of the sequential prisoners’ dilemma. Regarding second-mover behavior in the sequential prisoners’ dilemma, the model had predictive power at the individual level but not at the aggregate level. In addition to our analysis based on the point estimates of the inequality-aversion parameters, we checked more broadly for correlations across the decisions of the

³³On a related matter, if subjects exhibit concerns for efficiency or surplus maximization (see e.g. Charness and Rabin, 2002; Engelmann and Strobel, 2004), this may have biased our parameter estimates and also influence other choices. This does not, however, affect our hypotheses with respect to the correlations with the inequality aversion parameters at the individual level as long as efficiency concerns are not correlated with inequality aversion in some systematic way.

experiment. It turns out that the Fehr and Schmidt (1999) model predicts several of these correlations correctly, particularly some of the decisions associated with reciprocity, but other predicted correlations do not materialize. Therefore, we conclude that the model does not perform well at the individual level and that the aggregate support of the theory, if remarkable, should not be equated to individual-level validity.

We believe that the success of the inequality aversion model at the aggregate level could be based on an ability to qualitatively capture different important motives in different games but that the low predictive power of the model at an individual level is driven by the low correlation of these motives within subjects. Thus it appears to be both the strength and the weakness of the inequality aversion model that it can capture different motives in one functional form. On the one hand, this permits several apparently disparate results to be rationalized in one simple model. On the other hand, an individual's behavior is not well captured by the same model as different motives drive behavior in different situations and this is not reflected by the model. The inequality aversion model of Fehr and Schmidt (1999) can hence serve as an elegant "as if" model in several situations one at a time, but it does not appear to accurately and consistently reflect the preferences of individuals.

There are examples in the literature where a theory predicts the aggregate level well but fails at the individual level. Well known studies include market entry games where the standard Nash equilibrium works surprisingly well at the aggregate level but where no support is found at the individual level (e.g., Rapoport and Erev, 1998). Kahneman (1988) writes that the market entry games work "like magic". Another example are posted-offer markets with a mixed-strategy Nash equilibrium. The distribution of prices is approximated reasonably well by the prediction in such markets, even though individual pricing patterns are clearly inconsistent with the mixed-strategy Nash equilibrium (Davis and Wilson, 1998).

Generally, the aggregate support of a model in experiments constitutes a remarkable success of economic theory. How important failure at the individual level is may depend

on the interest of the researcher. Some researchers may find the individual-level failure of a theory intriguing and as a motive to search for further explanations of individual behavioral patterns, others may be perfectly content if a theory rationalizes the data at the aggregate level. Following Friedman (1953), the failure of a model at the individual level could be discarded as analytically irrelevant as long as aggregate results are broadly correct. However, whereas Friedman (1953) and most standard economics emphatically deny the descriptive accuracy of its behavioral assumptions, the other-regarding preferences models are explicitly descriptive behavioral theories (e.g., Fehr and Schmidt, 2006). Whether these models are “as if” approximations or indeed realistic descriptive models of individual behavior seems crucial here.

Finally, we would like to concede that the within-subjects test we have applied to the Fehr and Schmidt (1999) model is possibly a very demanding one. Little is known about how subjects play across games as individual-level comparisons have only rarely been conducted.³⁴ The main reasons for focussing on the Fehr and Schmidt model here were practical ones and the success it has achieved in the past. If we conclude that this model performs poorly at the individual level, then this finding is subject to the disclaimer that we do not know how other theories perform across different games. We believe that more research is needed with respect to both tests of other models and tests across other games.

³⁴Friedman and Sunder (2004) review empirical evidence on risk preferences. They argue that researchers have only rarely tried to compare individual decisions across different contexts. Moreover, even when such attempts have been made, risk preferences inferred from one experiment were often not able to account for other decisions. See, however, Andersen et al. (2005) who report stable risk preferences when subjects had to repeat the same risk-aversion task at two points in time.

References

- [1] Andersen, S., Harrison, G.W., Lau, M.I. and Rutström, E.E. (2005). “Lost in State Space: Are Preferences Stable?” Mimeo.
- [2] Andreoni, J. (1995). “Cooperation in Public Goods Experiments: Kindness or Confusion?” *American Economic Review* 85, 891-904.
- [3] Andreoni, J., Castillo, M. and Petrie, R. (2003). “What do Bargainers’ Preferences Look Like? Exploring a Convex Ultimatum Game.” *American Economic Review* 93, 672-685.
- [4] Andreoni, J. and Miller, J.H. (2002). “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism.” *Econometrica* 70, 737-753.
- [5] Berg, J., Dickhaut, J. and McCabe, K. (1995). “Trust, Reciprocity and Social History.” *Games and Economic Behavior* 10, 122-142.
- [6] Bolton, G.E. (1991). “A Comparative Model of Bargaining: Theory and Evidence.” *American Economic Review* 81, 1096-1136.
- [7] Bolton, G.E. and Ockenfels A. (2000). “ERC: A theory of Equity, Reciprocity and Competition.” *American Economic Review* 90, 166-193.
- [8] Camerer, C.F. (2003). “Behavioral Game Theory: Experiments in Strategic Interaction.” Princeton: Princeton University Press.
- [9] Charness, G. and Rabin, M. (2002). “Understanding Social Preferences with Simple Tests.” *Quarterly Journal of Economics* 117, 817-869.
- [10] Clark K. and Sefton, M. (2001). “The Sequential Prisoner’s Dilemma: Evidence on Reciprocation.” *The Economic Journal* 111, 51- 68.
- [11] Cox J. C., Friedman, D. and Gjerstad, S. (2006). “A Tractable Model of Reciprocity and Fairness.” *Games and Economic Behavior*, forthcoming.
- [12] Davis, D.D. and Wilson, B. (1998). “Mixed Strategy Nash Equilibrium Predictions as a Means of Organizing Behavior in Posted-Offer Market Experiments.” In: C. Plott and Smith, V.L. (Eds.): “Handbook of Experimental Economics Results.” Amsterdam: North-Holland, forthcoming.
- [13] Dawes, R.M. (1989). “Statistical Criteria for Establishing a Truly False Consensus Effect.” *Journal of Experimental Social Psychology* 25, 1-17.
- [14] Engelmann, D. and Strobel, M. (2000). “The False Consensus Effect Disappears if Representative Information and Monetary Incentives Are Given.” *Experimental Economics* 3, 241-260.

- [15] Engelmann, D. and Strobel, M. (2004). "Inequality Aversion, Efficiency and Maximin Preferences in Simple Distribution Experiments." *American Economic Review* 94, 857-869.
- [16] Falk, A. and Fischbacher, U. (2006) "A Theory of Reciprocity." *Games and Economic Behavior* 54, 293-316.
- [17] Fehr, E., Kirchsteiger, G. and Riedl, A. (1993). "Does Fairness prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics* 108, 437-460.
- [18] Fehr, E. and Schmidt, K.M. (1999). "A theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics* 114, 817- 868.
- [19] Fehr, E. and Schmidt, K.M. (2005). "The Rhetoric of Inequity Aversion – A Reply." Mimeo
- [20] Fehr, E. and Schmidt, K.M. (2006). "The Economics of Fairness, Reciprocity, and Altruism—Experimental Evidence and New Theories." In: Kolm, S. and Ythier, J.M. (Eds.): "Handbook on the Economics of Giving, Reciprocity and Altruism." Amsterdam: North Holland, forthcoming.
- [21] Fischbacher, U. (1999). "z-Tree - Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual." Working Paper Nr. 21, Institute for Empirical Research in Economics, University of Zurich.
- [22] Forsythe R., Horowitz J.L., Savin N.E. and Sefton M. (1994). "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior* 6, 347-369.
- [23] Friedman, D. and Sunder, S. (2004). "Risky Curves: From Unobservable Utility to Observable Opportunity Sets." Mimeo.
- [24] Friedman, M. (1953). "The Methodology of Positive Economics" in Friedman, M.: "Essays in Positive Economics." Chicago: University of Chicago Press.
- [25] Goeree, J., Holt, C.A., and Laury, S. (2002). "Private Costs and Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior." *Journal of Public Economics* 83, 257-278.
- [26] Güth, W.Schmittberger, R. and Schwarze B. (1982). "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3, 367-388.
- [27] Harrison, G. (1989). "Theory and Misbehavior of First-Price Auctions." *American Economic Review* 79, 749-762.
- [28] Holt, C.A. and Laury, S. (2002). "Risk Aversion and Incentive Effects." *American Economic Review* 92, 1644-1655.

- [29] Huck, S., Müller, W. and Normann, H.T. (2001). "Stackelberg beats Cournot—On Collusion and Efficiency in Experimental Markets." *Economic Journal* 111, 749-765.
- [30] Kahneman, D. (1988). "Experimental Economics: A Psychological Perspective." In: R. Tietz, W. Albers and R. Selten (Eds.): "Bounded Rational Behavior in Experimental Games and Markets." Berlin-Heidelberg-New York: Springer.
- [31] Kahneman, D., Knetsch, J. and Thaler, R. (1986). "Fairness and the Assumptions of Economics." *Journal of Business* 59, S285-S300.
- [32] Ledyard, J.O. (1995). "Public Goods: A Survey of Experimental Research." In: Kagel, J.H. and Roth, A.E. (Eds.): "The Handbook of Experimental Economics." Princeton: Princeton University Press.
- [33] Levine, D. (1998). "Modelling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics* 1, 593-622.
- [34] McKelvey, R.D. and Palfrey, T.R. (1998). "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics* 1, 9-41.
- [35] Mullen, B., Atkins, J.L., Champion, D.S., Edwards, C., Hardy, D., Story, J.E. and Venderklok, M. (1985). "The False Consensus Effect: A Meta-Analysis of 115 Hypothesis Tests." *Journal of Experimental Social Psychology* 21, 263-283.
- [36] Oosterbeek, H., Sloof, R. and van de Kuilen, G. (2004). "Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis." *Experimental Economics* 7, 171-188.
- [37] Rapoport, A., and Erev, I. (1998). "Coordination, 'Magic', and Reinforcement Learning in a Market Entry Game." *Games and Economic Behavior* 23, 146-175.
- [38] Ross, L., and Nisbett, R.E. (1991). "The Person and the Situation. Perspectives of social psychology." New York: McGraw-Hill.
- [39] Roth A.E. (1995). "Bargaining Experiments." In: Kagel, J.H. and Roth, A.E. (Eds.): "The Handbook of Experimental Economics." Princeton: Princeton University Press.
- [40] Shaked, A. (2005). "The Rhetoric of Inequity Aversion." Mimeo.

Appendix

Proofs

Here, we formally derive the hypotheses of the results section. Some proofs can also be found in F&S.

Hypothesis 1 (i) *Subjects with $\beta_i > 0.5$ should offer $s_i = 10$ in the Ultimatum Game.* (ii) *Subjects with $\beta_i < 0.5$ may, depending on their beliefs, offer either $s_i = 10$ or $s_i < 10$ in the Ultimatum Game.*

Proof. An offer of $s = 10$ will surely be accepted by all responders and thus gives the proposer a utility of $U_i(10, 10) = 10$. Offering $s < 10$ either gives zero utility to the proposer if the offer is rejected or $U_i(20 - s, s) = 20 - s - \beta_i(20 - 2s)$ if it is accepted. When $\beta_i > 0.5$, we have $20 - s - \beta_i(20 - 2s) < 10$, hence, these subjects will choose $s = 10$. When $\beta_i < 0.5$, by contrast, $20 - s - \beta_i(20 - 2s) > 10$ and the proposer gains from offering $s < 10$ if the offer is accepted. Whether or not a subject with $\beta_i < 0.5$ will actually offer $s < 10$ depends on the beliefs whether such an offer will be accepted. ■

In the next hypothesis, let y_i denote the contribution of subject i in the PG.

Hypothesis 2 (i) *Subjects with $\beta_i < 0.3$ should choose $y_i = 0$ in the PG.* (ii) *Subjects with $\beta_i > 0.3$ may, depending on their beliefs, contribute any $y_i \in [0, 10]$ in the PG.*

Proof. Suppose player i believes that player j will contribute $\bar{y} \in [0, 10]$ so that the payoff for player i is $10 - y_i + 0.7(y_i + \bar{y}) = 10 + 0.7\bar{y} - 0.3y_i$ and the payoff of player j is $10 + 0.7y_i - 0.3\bar{y}$. If player i also contributes \bar{y} , he gets a utility of $10 + 0.4\bar{y}$. If player i contributes $y_i < \bar{y}$, this yields a utility of $10 + 0.3(\bar{y} - y_i) + 0.4\bar{y} - \beta_i(\bar{y} - y_i)$ which is larger than $10 + 0.4\bar{y}$ if, and only, if $\beta_i < 0.3$. If player i contributes $y_i > \bar{y}$, this yields a utility of $10 - 0.3(y_i - \bar{y}) + 0.4\bar{y} - \alpha_i(y_i - \bar{y}) < 10 + 0.4\bar{y}$ if $\alpha \geq 0$, which holds for every subject. Hence, player i will never contribute more than \bar{y} , will, depending on his beliefs, contribute $\bar{y} \in [0, 10]$ if $\beta_i > 0.3$, and will contribute $y_i = 0$ if $\beta_i < 0.3$. Note, finally, as players with $\beta_i < 0.3$ should choose $y_i = 0$ for any degenerate belief, they should also choose $y_i = 0$ for any non-degenerate belief on y_j . ■

Hypothesis 3 (i) *Given first-mover cooperation, second movers in the SPD should defect if, and only if, $\beta < 0.3$.* (ii) *Given first-mover defection, second movers in the SPD should defect.*

Proof. (i) If the first mover cooperates, player i prefers to defect if, and only, if $U_i(14, 14) < U_i(17, 7)$, that is, if, and only, if, $14 < 17 - \beta_i(17 - 7) \Leftrightarrow \beta_i < 0.3$. (ii) If the first mover defects, player i is better off defecting regardless of the inequality parameters since $U_i(10, 10) = 10 > U_i(7, 17) = 7 - 10\alpha_i$ and $\alpha_i \geq 0$. ■

Hypothesis 4 *If subjects know the true distribution of the β parameter, first-movers in the SPD should cooperate if, and only if, $\alpha_i < 0.52$.*

Proof. If the first mover defects, the second mover will also defect (regardless of α_j and β_j) and both players get $U_i(10, 10) = 10$. Let the first mover's belief for the second mover to cooperate be p . Then the expected payoff from cooperating is $pU_i(14, 14) + (1 - p)U_i(7, 17)$, and cooperating yields an expected payoff higher than defecting if, and only if,

$$\alpha_i < \tilde{\alpha} = \frac{7p - 3}{10(1 - p)}.$$

From the analysis of the second movers above, we know that second movers reciprocate cooperation if, and only if, $\beta_i > 0.3$. In the data, we have 41 subjects with $\beta_i > 0.3$. Hence, $p = 41/61 = 0.672$. Using this value of p , cooperating as a first mover pays if, and only if, $\alpha_i < 0.52$. ■

Characterization of the MDG (not intended for publication)

The purpose of the MDG is to obtain a (near) point estimate of the β parameter for rational F&S-type of players with $\beta_i \in [0, 1)$. In this appendix, we show that the MDG design we use is the simplest design to obtain such an estimate in an environment uncontaminated by intentions and beliefs.

Such an estimate of the β parameter can be found if, and only if, we can elicit the point where player i is indifferent between two outcomes (x_i, x_j) and (x'_i, x'_j) such that

$$x_i - \beta_i(x_i - x_j) = x'_i - \beta_i(x'_i - x'_j). \quad (4)$$

For this equality to have a unique solution in β_i , we need to impose three conditions here. First, we need $x_i \geq x_j$ and $x'_i \geq x'_j$ with at least one inequality being strict—otherwise the β parameter would not apply at all. Second, we do not get any information from the trivial solution where $(x_i, x_j) = (x'_i, x'_j)$. Third, we need $\text{sign}(x_i - x'_i) = \text{sign}(x_i - x_j - (x'_i - x'_j))$ because otherwise one outcome is strictly preferred to the other for any β_i . Without loss of generality, we can set $x_i = x_j$ and obtain

$$x_i = x'_i - \beta_i(x'_i - x'_j) \quad (5)$$

or

$$\beta_i = \frac{x'_i - x_i}{x'_i - x'_j} \quad (6)$$

We want to get a (near) point estimate through binary choices. So we need to let subjects make choices between various outcomes (corresponding to one side of (5)) and a constant outcome (corresponding to the other side of (5)). The choices must be designed

such that any player with $\beta_i \in [0, 1)$ will prefer x_i over $x'_i - \beta_i(x'_i - x'_j)$ for at least one but not for all binary choices of the game. In that case, we know that player i has some $\beta_i \in [\underline{\beta}, \overline{\beta}]$ with $0 \leq \underline{\beta} \leq \beta_i \leq \overline{\beta} < 1$.

For our MDG, we decided to keep the right-hand side of (5) constant (with $x'_i = 20$ and $x'_j = 0$) and vary the left-hand side (with $x_i \in \{0, 1, 2 \dots 20\}$). Now, all players with $\beta_i \in [0, 1)$ prefer (20,0) over (0,0) and they also (weakly) prefer (20,20) over (20,0). It follows that our MDG is suitable to elicit the β_i parameter. In particular, it also allows us to detect whether there are any subjects with $\beta_i \geq 1$, namely if they choose (0,0) over (20,0).

Consider the alternative to keep the left-hand side constant and vary the right-hand side. We obviously need only consider $x'_i \geq x_i$ and $x'_j \leq x_i$. Let us first keep $x'_i > x_i$ fixed. By varying x'_j between 0 and x_i , we can detect any β between $(x'_i - x_i)/x'_i$ and 1. If, however, a subject prefers $(x'_i, 0)$ over (x_i, x_i) we can only conclude that $\beta_i \leq (x'_i - x_i)/x'_i$, where $(x'_i - x_i)/x'_i > 0$ by assumption. (Even if we allow the rather unrealistic case of $x'_j < 0$, this problem does not disappear since x'_j will obviously have to be finite. Furthermore, if we choose $x_i > x_j$, the denominator of β_i will be $x'_j - (x_i - x_j)$, and hence the minimal β_i that could be detected would increase.) In order to detect whether there are subjects with $\beta_i = 0$, we need to add another choice where $x'_i = x_i$ and $x'_j < x_i$, because all subjects with $\beta_i > 0$ will prefer (x_i, x_i) over (x_i, x'_j) . Hence in order to investigate the whole interval $[0, 1]$, we need to vary both x'_i and x'_j across choices, which is arguably more complicated for subjects than our design.

Alternatively, let us keep $x'_j < x_i$ fixed. By varying x'_i between x_i and $x_i + k$, we can identify all β_i between 0 and $k/(k + x_i - x'_j)$. If a subject prefers (x_i, x_i) over $(x_i + k, x'_j)$, we can only conclude that $\beta_i \geq k/(k + x_i - x'_j)$, where $k/(k + x_i - x'_j) < 1$. (If we choose $x_i > x_j$, the denominator of β_i will be $k + (x_i - x'_j) - (x_i - x_j)$. While this increases the maximal β that could be identified, it will still be smaller than 1 since $(x_i - x'_j) > (x_i - x_j)$, because in order to detect any β_i smaller than 1, the fixed x'_j has to be smaller than x_j .) Since k obviously has to be kept finite, in order to detect whether there are subjects with $\beta_i \geq 1$, we have to add another choice where $x'_i > x_i$ and $x'_j = x_i$ because all subjects with $\beta_i < 1$ will prefer (x'_i, x_i) over (x_i, x_i) . Hence again we would have to vary both x'_i and x'_j across choices in order to study the whole range of permissible β . Consequently, our design (except setting $x_i = x_j$, which is no restriction) is structurally the simplest design to provide a (near) point estimate for the whole range of relevant β .

Tables

Game	label	description
ultimatum game	UG	£20 pie, proposer gets £(20-s) and responder s if the responder accepts, both get zero otherwise
modified dictator game	MDG	dictator chooses between £20-£0 and equitable outcomes ranging from £0-£0 to £20-£20
sequential prisoners' dilemma	SPD	both defect: £10-£10; both cooperate: £14-£14; one defects, one cooperates: £17-£7
public good game	PG	two players, £10 endowment per player, marginal per capita return on contributions is 0.7

Table 1. The experimental design. Each subject played all four games once. In the UG and the SPD, subjects made decisions at all nodes. All choices were made without feedback on decisions of earlier games. Therefore, the strategy-elicitation method was used where necessary.

α	F&S	data	β	F&S	data
$0 \leq \alpha < 0.4$	30%	31%	$0 \leq \beta < 0.235$	30%	29%
$0.4 \leq \alpha < 0.92$	30%	33%	$0.235 \leq \beta < 0.5$	30%	15%
$0.92 \leq \alpha < 4.5$	30%	23%	$0.5 \leq \beta \leq 1$	40%	56%
$4.5 \leq \alpha < \infty$	10%	13%			

Table 2. Distribution of α and β as assumed in F&S and as observed in our data.

	α	β	UG offer	PG	SPD 1 st	SPD 2 nd
α	—	-0.03	0.40**	0.07	-0.03	0.19
β		—	0.13	0.13	0.04	0.34**
UG offer			—	0.19	0.13	0.49**
PG				—	0.24	0.41**
SPD 1 st					—	0.43**
SPD 2 nd						—

Table 3. Correlations between decisions (Spearman ρ for two ordinally scaled variables; rank biserial correlation for one ordinally scaled variable and one dichotomous nominally scaled variable; phi coefficient for two dichotomous nominally scaled variables), ** indicates significance at the 1% level.

Figures

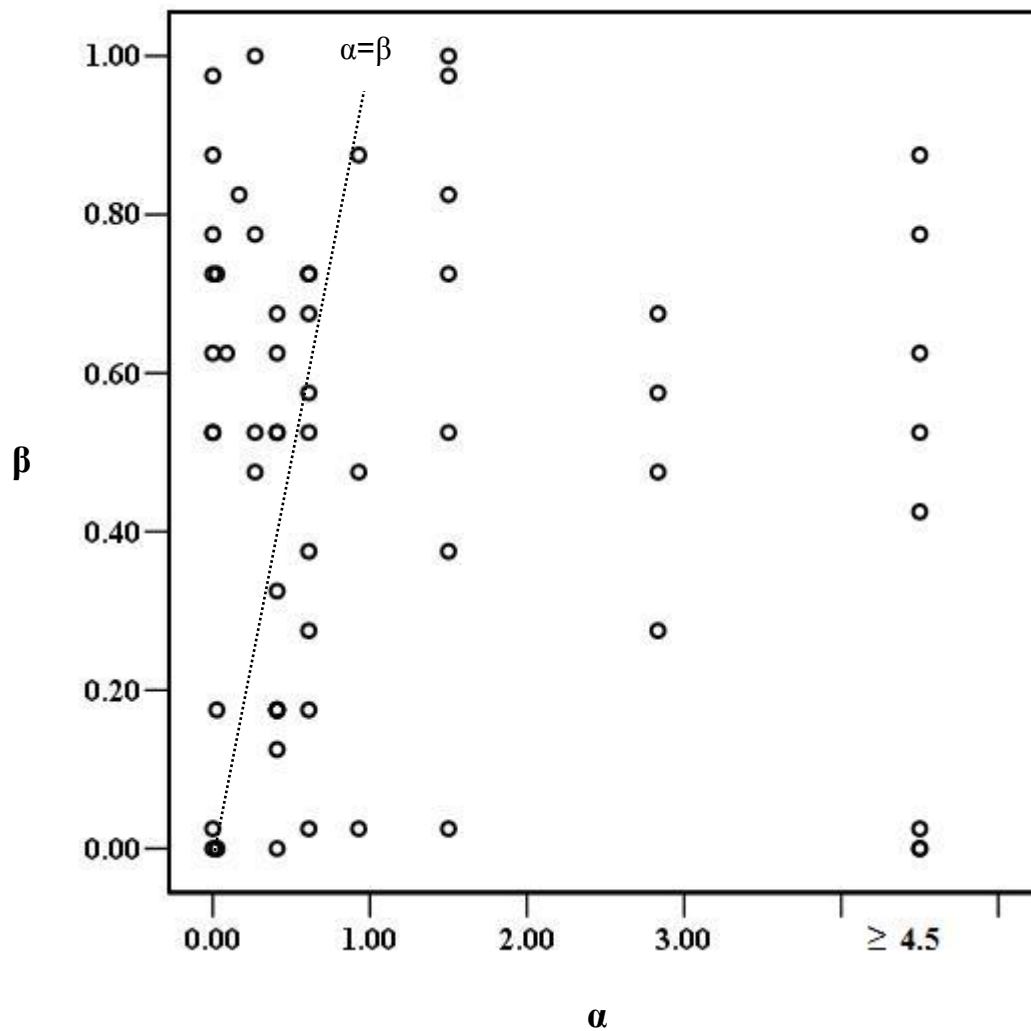


Figure 1: The joint α - β distribution. Each dot in the figure represents an individual's α and β parameter. Observations to the left of the $\alpha = \beta$ line have $\alpha < \beta$. Observations with the highest level of α cannot be pinned down more narrowly than $\alpha \geq 4.5$.

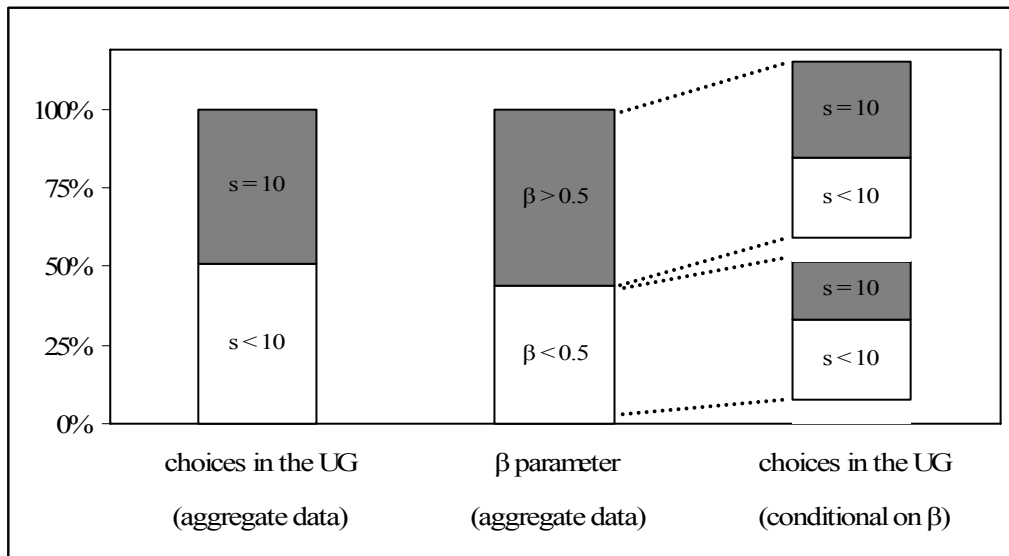


Figure 2: Aggregate versus individual-level analysis of UG offers. The left and the middle column show the proportions of $s=10$ (equal split) offers in the UG and the share of $\beta > 0.5$ subjects, respectively, at the aggregate level. These proportions are roughly equal which is consistent with F&S at the aggregate level. The right column shows UG offers conditional on the individual β parameters. Subjects with $\beta > 0.5$ should offer $s=10$ but only slightly more than half of them (55%) do. The F&S theory is therefore rejected at the individual level.

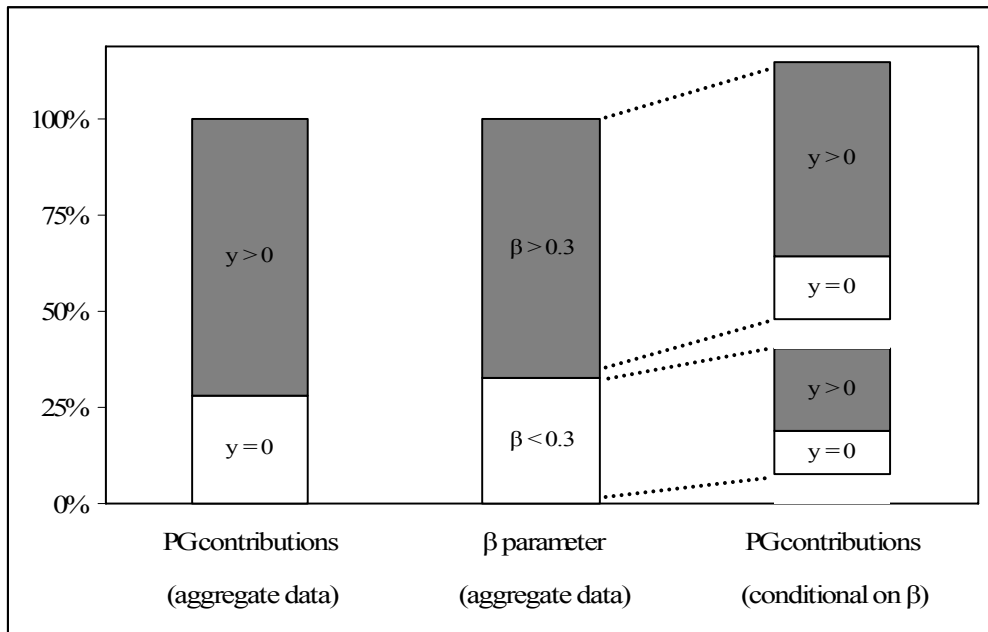


Figure 3: Aggregate versus individual-level analysis of PG contributions (denoted by y). The left column shows the proportion of zero contributions and the middle column the share of $\beta < 0.3$ subjects. The proportions are rather equal which is consistent with the F&S theory at the aggregate level. The right column shows PG contributions conditional on the individual β parameters. Subjects with $\beta < 0.3$ should not contribute but almost two thirds of them do. The theory is therefore rejected at the individual level.

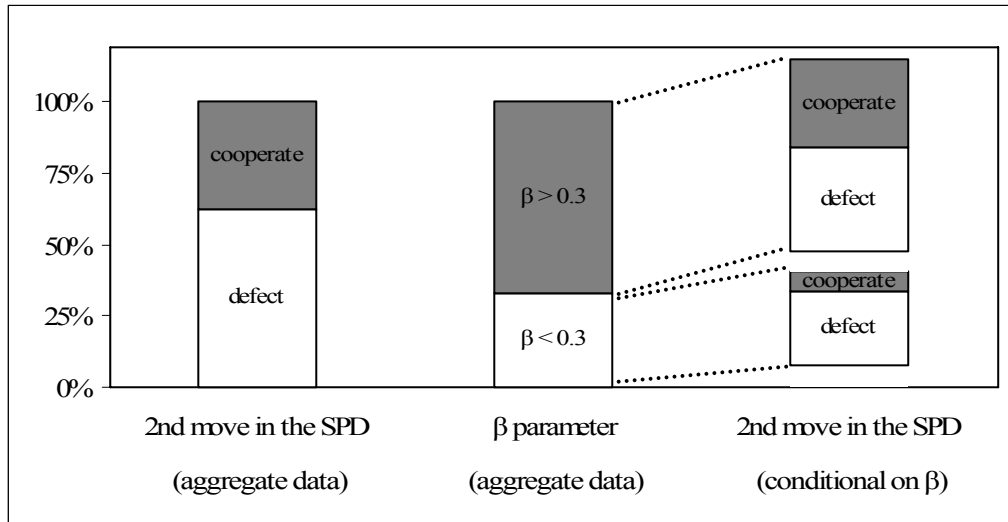


Figure 4: Aggregate versus individual-level analysis of the second move in the SPD. The left and the middle column show the proportions of cooperate choices in the SPD and the share of $\beta > 0.3$ subjects, respectively. The F&S theory predicts that subjects should cooperate if and only if $\beta > 0.3$. As these proportions differ at the aggregate level, the theory is rejected. Looking at cooperation decisions conditional on β (right column), a larger share of the subjects defects when $\beta < 0.3$, providing support of the F&S theory at the individual level.

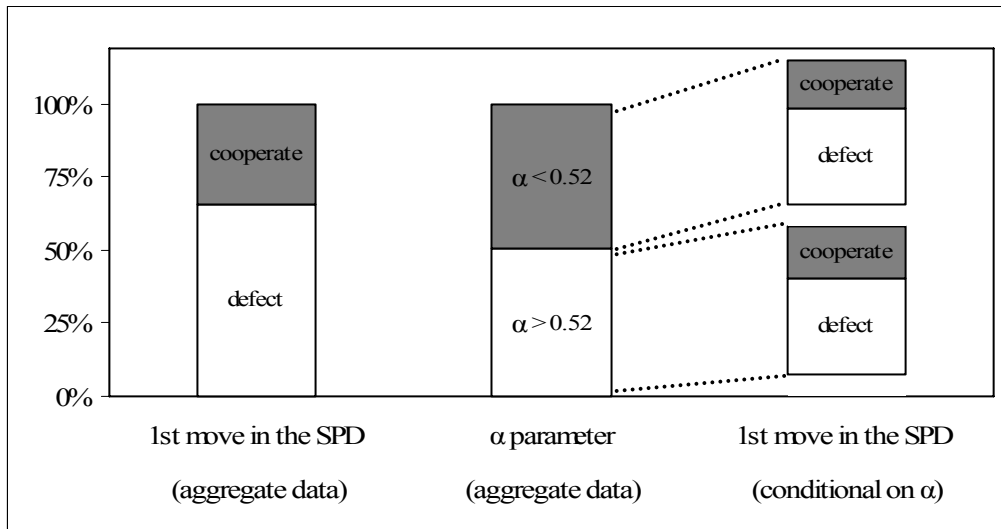


Figure 5: Aggregate versus individual-level analysis of the first move in the SPD. The left and the middle column show the proportions of cooperate choices in the SPD and the share of $\alpha < 0.52$ subjects, respectively. The F&S theory predicts that subjects should cooperate if and only if $\alpha < 0.52$, provided subjects know the distribution of the β parameter. The differences between these proportions do not differ significantly at the aggregate level which supports the theory. Looking at cooperation decisions conditional on α (right column), the proportions of cooperators and defectors are virtually identical regardless of the α parameter, rejecting the F&S theory at the individual level.