A Within-Subject Analysis of Other-Regarding Preferences*

Mariana Blanco† Dirk Engelmann‡ Hans-Theo Normann§

Royal Holloway, University of London

PRELIMINARY VERSION

April 28, 2006

Abstract

In this paper we assess the predictive power of inequality aversion using a within-subjects design. We run four different experiments (an ultimatum game, a dictator game, a sequential prisoner's dilemma and a public-good game) with the same sample of experimental subjects. This design allows us to make within-subjects comparisons across different games. We use the responder data from the ultimatum game in order to estimate a parameter of aversion against disadvantageous inequality, and we take data from a modified dictator game to estimate a parameter against advantageous inequality. We then use this joint distribution to test several hypotheses about individual behavior in the other games. Our results show that the inequality aversion model has some predictive power at the aggregate level but fails almost entirely at the individual level.

JEL Classification numbers: C72, C91.

Keywords: Other-regarding preferences, inequality aversion, experimental economics

*We are grateful to Nikos Nikiforakis, Juan Vargas as well as seminar and conference audiences at Ben-Gurion University, Economic Science Association (ESA) Allessandria, Exeter University, Royal Economic Society (RES) Nottingham and Royal Holloway College for helpful comments.

†Department of Economics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK, Fax: +44 1784 439534, e-mail: *m.blanco*@rhul.ac.uk.

‡Department of Economics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK, Fax: +44 1784 439534, e-mail: *dirk.engelmann*@rhul.ac.uk.

§Department of Economics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK, Fax: +44 1784 439534, e-mail: *hans.normann@rhul.ac.uk*.

1 Introduction

Behavioral economists are currently investing substantial efforts into developing models of social or other-regarding preferences. The starting point of this literature is that data from, for example, the ultimatum game (Güth et al., 1982), the dictator game (Kahneman et al., 1986; Forsythe et al., 1994) and the trust game (Berg et al., 1995) are by and large inconsistent with the economics paradigm of self-interested utility maximization. By relaxing the assumptions of the standard model and allowing for other-regarding motives, the behavioral models attempt to consistently explain these experimental results.

Models of other-regarding preferences take into account that behavior is usually heterogenous among participants. A robust finding of various experimental studies is that not all participants behave alike. Not everybody behaves perfectly selfish as the traditional model asserts but not all subjects behave perfectly, say, altruistic either. The heterogeneity of behavior among subjects implies that models of other-regarding preferences typically make predictions about the distribution of behavior. In particular, they predict how this distribution is expected to stay constant (or vary) across two games. For example, based on the properties of two fictitious games A and B, an other-regarding preferences model could imply the hypothesis that "subjects should be equally likely to make the altruistic choice in game A and game B". The experimental evidence in favor or against a behavioral model depends on whether or not such comparisons of behavior across games are in line with the model's predictions.

The starting point of this paper is that predictions of this kind can be tested in two ways. To the best of our knowledge, the existing literature has almost exclusively relied on *aggregate-level* or unrelated-sample experimental data. In the above example, the hypothesis is simply tested by comparing the share of altruistically behaving subjects in game A and B when the data of the two games come from different subject samples. Only if there is a significantly higher share of altruistic choices in one game the hypothesis is rejected. However, the same hypothesis can also be tested with *individual-level* or related-sample data. In a related-sample of game A and game B decisions, the experimenter can analyze each individual's behavior in the two games with respect to the hypothesis. The within-subjects test would be whether it is actually the same subjects who make the altruistic choice in both game A and B, as predicted by other-regarding preferences theories.

We believe that such individual-level data tests of these theories are crucial when it comes to their behavioral validity. All such models we are aware of explicitly make individual-level predictions. They are emphatically models of individual behavior, making allegedly realistic behavioral assumptions. The other-regarding preferences models are generally considered to by and large correctly predict aggregate outcomes across several games. While this constitutes remarkable progress in the interpretation of recent experimental findings, we believe this predictive success is incomplete without individual-level data support.

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) explicitly 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 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 is not the case with the experiments that motivated the other-regarding preferences literature.

The main novelty of this paper is to provide a first systematic[1] 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 later refined by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). Basically, inequality aversion stipulates that individuals do not only care for their own material payoff but also for the distribution of payoffs among players. In particular, individuals dislike both having a lower as well as a higher payoff as others, and so, all else equal, an equal distribution of payoff maximizes their utility. Judging from citation figures, the inequality model is very popular among experimental economists. Even though the model fails in some experiments, the more recent extended behavioral models that aim

[1] Andreoni et al. (2003) do 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 role), and the individual-level comparison of decisions is only an aside of the analysis. Charness and Rabin (2002) conduct a related-sample experiment where subjects played two or four different games. They explicitly mention that they do not conduct an individual-level data analysis.

at explaining results in these experiments nevertheless include some concerns for equality.[2]

In the main part of 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 classical 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 towards disadvantageous inequality, and we take data from the modified dictator game to estimate a parameter of aversion towards 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 individual behavior in the other games.

Our results show that the model has considerable predictive power at the aggregate level but fails almost entirely at the individual level. With one exception, the model does not predict the right correlation of decisions. That is, the degree of inequality aversion that an individual exhibits in the ultimatum and modified dictator games 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 exprimental 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 in Section 7 we conclude.

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

[2] For example, Falk and Fischbacher (2006) explicitly builds on Fehr and Schmidt (1999), while Charness and Rabin (2002) and Cox, Friedman and Gjerstad (2004) assume that subjects are more altruistic towards others with relatively low payoffs.

subject made a decision in both roles.[3] 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 the participants in the session 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 done without any information on other subjects' choices and without any communication. At the end of the experiment, one game was chosen randomly and subjects were randomly matched in pairs. In all games except for the public-good game, 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 randomly chosen game. 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 decisive pair of decisions. We believe that our design is appropriate to minimize confounding effects between games and to avoid that subjects average their earnings across games.[4] Regarding feedback and payments, our design is very similar to the one in Charness and Rabin (2002).[5]

When selecting the games for our experiment, we wanted to make sure that we include the ones most relevant in the other-regarding preferences literature. Therefore, we chose the ultimatum game, the dictator

[3] The dictator game involves a decision maker and a passive player. Each subject made a choice in the role of the decision maker.

[4] Alternatively, we could have invited subjects to four separate sessions, with some time gap between the sessions. With such a design, inevitably, some subjects would not show up for some sessions causing the data to be incomplete. More importantly, the emotional state on each of these separate sessions may differ and confound the data.

[5] See their paper for a further discussion of issues arising due to the related-sample design.

game, and the public-good game. The trust game is also rather frequently discussed in the literature. Instead of the trust game, we used a sequential prisoner's dilemma. This is essentially a simplified trust game reduced to two decisions for each player. It has the same qualitative properties as the trust game (in terms of trust and trustworthiness) but is much simpler and therefore more suitable for our purposes.[6]

We also had to decide on the number of games to be played. Four games seemed a reasonable compromise to us between generating a rich data set and maintaining salient incentives (Smith, 2002). With a higher number of games and decisions, we might have risked that subjects did not care about each individual decision any more. Charness and Rabin (2002) also ran a maximum of four games per subjects in their experiments.

We now introduce the four games as implemented in our experiment. 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 (£z) to the responder, keeping £20− £z 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− £z and £z, respectively. As mentioned above, we let 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. That is, there were 21 different distributions to decide upon.[7] 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 on 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 is not suitable to yield a point prediction of the parameter

[6] Conducting a standard trust game would either require to give feedback, which would violate our design approach, or to use the strategy method for second movers, which would make the game rather complex because the action space of second movers depends on the first move.

[7] Correspondingly, the proposer's offers were restricted to integers.

measuring aversion against advantageous inequality.[8] Therefore, we implemented a modified dictator game (henceforth MDG) more resembling the orginal 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 of (£x, £x) where $x \in \{0, 1..., 20\}$.[9] 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 player 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 respective decisions.

Finally, the public-good game (henceforth PG) we used was a simple two-player voluntary contribution

[8] Following the formulation of F&S (see below), dictators in the standard game keep the whole amount if $\beta < 0.5$ and split it equally if $\beta > 0.5$ which hence allows only a very coarse classification of subjects based on the standard dictator game.

[9] In this modified dictator game, a purely self-centered individual would always choose (£20, £0) over (£x, £x) where $x \in \{0, 1..., 19\}$ and would be indifferent between (£20, £0) and (£20, £20). A dictator strongly disliking advantageous inequality, would choose the right column, no matter how small x. Subjects between these two extremes would be expected to switch from choosing the left column to the right column. The reason why subjects with monotone preferences should switch at some some point x^* from choosing the left columns to the right column (if at all), is that the egalitarian outcome is "cheaper" for all decisions $x > x^*$. Hence, if a player prefers (£x^*, £x^*) to (£20, £0), this player should also prefer (£$x^* + 1$, £$x^* + 1$) to (£20, £0).

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 per capita rate). Note that, when restricting actions to the extreme choices of zero and full contribution, the payoffs are 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.

We implemented two different sequences in which the games were played. Because of the similarity of the games, we wanted to avoid that either UG and MDG or PG and SPD are played in a row. Also, because of length of the instructions and a control questionnaire, we wanted the public-good 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.[10] 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.[11] In the data analysis below, we discarded 11 out 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

[10] Given that the ultimatum game and the sequential PD can be played in two and six sequences, respectively, we would even have to take 288 different variants into account. See also Charness and Rabin (2002) who do not control for sequencing effects. Andreoni et al. (2003) do not control for alternative sequences either.

[11] See Fehr, Naef, and Schmidt (2006) and Engelmann and Strobel (2006) for a discussion whether economics majors may behave differently in distribution experiments.

analysis. Henceforth, we will deal with a total of 61 subjects.[12] The experimental software was developed in z-Tree (Fischbacher, 1999). Sessions lasted about 50 minutes and the average earnings were £11.

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[13] 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 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, we follow the categorization in F&S. All subjects were willing to accept the equal split. About 20% of our subjects accepted only more than 4/9 of the pie. Another 16% had acceptance thresholds between 1/3 and 4/9. About 33% had their acceptance level between 1/4 and 1/3 and the remaining 31% accepted less than 1/4. These figures differ only in the top segment from the ones in F&S which were derived from data in Roth (1995). The distribution they report has 10% of the subjects accepting only more than 4/9 and 30% for each of the other three segments (we formally test for differences between these distributions below).

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% do not switch at all, that is, they even prefer (£20, £0) over (£20, £20). Two out of 61 subjects prefer (£0, £0) over (£20, £0). Because we modified the dictator game,

[12] Holt 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, and they also decided to drop some subjects from their analysis for this reason.

[13] Our terminology follows Andreoni et al. (2003) here.

our version has no precedent in the literature and the results cannot be compared to those reported for standard DG experiments. One parallel that can, however, be drawn is that Forsythe et al. (1994) found 20% of the dictators choosing 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 Kahnemann et al. (1986), 76% of dictators prefer (10,10) over (18,2) which compares to 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 as in Kahnemann 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%) defected as well. Our results are very similar to the ones obtained by Clark and Sefton (2001) in their SPD. The figures they obtained ("baseline" treatment, last round)[14] 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 of the endowment was contributed by 41% of the subjects including 28% (of the total population) who contributed nothing. A contribution of zero 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. Holt, Goeree and Laury (2002) report on one-shot public-good games. They have one treatment with two players where the marginal per capita return is similar to our's (0.8).[15] 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 Holt, Goeree and Laury (2002) since the endowment was 25 tokens, again, the results are remarkably similar to those we observed

[14] Clark and Sefton (2001) repeat their SPD and report cooperation rates in the first and the last round. 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 to the 20% gain in the "baseline" treatment of Clark and Sefton (2001).

[15] Most of the treatments in Holt, Goeree and Laury (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 our experiments and in the standard PGG.

in the PG. Differences to 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 (and even in the subgames of the ultimatum game and the SPD) despite our related-sample design. Therefore, our design should be suitable for the individual-level test of the inequality model.

4 Model and estimation of the inequality aversion parameters

In F&S outcome-based theory, other-regarding preferences are modeled as inequality aversion. This means that players are not only concerned 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

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

where $i \neq j$. F&S make the following 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 experimentas, F&S further assume that $\beta_i < n/(n-1)$ for $n = 6$, hence $\beta_i < 0.8\overline{3}$ (p. 832), and that α and β are positively correlated (p.864).[16]

We follow F&S in deriving the distribution of the parameter for aversion towards disadvantageous inequality, α, from the UG responder decisions. Since we employ the strategy elicitation method, the rejection levels in the ultimatum game give us (near) point estimates of α_i for each individual. To see this, suppose z_i' is the lowest offer responder i is willing to accept, and, consequently $z_i' - 1$ is the highest offer i rejected (recall that choices had to be integers). We conclude that this responder is indifferent between accepting some offer $z_i \in [z_i' - 1, z_i']$ and getting a zero payoff from a rejection. Therefore, we have

[16] On this point, see Shaked (2005).

$U_i(z_i, 20 - z_i) = z_i - \alpha_i(20 - z_i - z_i) = 0$.[17, 18]

Thus, the estimate of the parameter of disadvantageous inequality is

$$\alpha_i = \frac{z_i}{2(10 - z_i)}. \tag{2}$$

In our data, we set $z_i = z_i' - 0.5$. This is somewhat arbitrary but it does in no way affect 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 $z_i' \leq 10$, so, division by zero cannot occur by assumption here. For a subject with $z_i' = 0$, we observe no rejected offer and we cannot infer the indifference point z_i. Therefore, we set $\alpha_i = 0$ for participants with $z_i' = 0$ 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.[19]

Let us now turn to the parameter of aversion towards advantageous inequality, β. F&S (1999) derive the distribution of this parameter from offers in the UG. There are various problems with this. First, proposers' offers depend on their beliefs about the other players' 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 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 purely selfish

[17] Note that any person who is averse to disadvantageous inequality will have a rejection threshold less than or equal to 10. Hence, as for measuring aversion towards disadvantageous inequality, only the range of offers up to half of the pie is relevant.

[18] Our UG design explicitly asks for acceptance or rejection of each possible offer. Interestingly, we observe seven subjects who consistently reject offers $z \geq z'$ for some $z' > 10$. Since $z > 10$ here, these decisions reveal also something about the subjects' β_i parameter. More precisely, such rejections imply $\beta_i > 1$. For these seven subjects, we find some relation to the β we estimate from MDG below. The average β_i of these subjects is significantly higher than that of the rest of the sample (Mann-Whitney U, $p = 0.058$), and for two of these subjects we will find $\beta \geq 1$ based on the MDG. In any event, responders could expect the probability to receive such an offer to be close to zero, such that their decisions are effectively cheap talk.

[19] See Charness and Rabin (2002) and Engelmann and Strobel (2004) for evidence that at least in non-strategic games such preferences are common.

proposer ($\beta = 0$) is to offer half of 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 analogously to the way the α were derived. 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 to provide a (near) point estimate for the whole range of relevant β.

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 (x_i, x_i) egalitarian distribution where $x_i \in [x_i' - 1, x_i']$ and $x_i' \in \{0, ..., 20\}$. From (1) we get $U_i(20, 0) = U_i(x_i, x_i)$ if and only if $20 - 20\beta_i = x_i$. This yields

$$\beta_i = 1 - \frac{x_i}{20}. \tag{3}$$

For our data analysis, we use $x_i = x_i' - 0.5$ (which, as above, does not affect our results), but this does not work well at the boundaries. Subjects with $x_i' = 0$ prefer $(0, 0)$ over $(20, 0)$, so they are possibly willing to sacrifice more than £1 in order to reduce the inequality by £1. Therefore, these subjects might have $\beta_i > 1$. However, 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.[20]

In Table 1, we report on the distribution of α and 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

[20] F&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).

(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 1, we prefer to interpret these mass points not literally but instead refer to the broader intervals which F&S derive on the way (see pp. 843-4).

The intervals in Table 1 for the α parameter correspond to those intervals F&S suggest for the rejection thresholds. 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 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.[21] 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).

As for the β distribution, we use the very intervals F&S (p. 844) derive. (F&S only assign the mass points within the intervals later on). 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). The analysis is further complicated by the fact that, as mentioned above, F&S impose $\beta_i < 0.8\overline{3}$ for all subjects in one of their proofs whereas we find 7 subjects (11%) with $\beta > 0.8\overline{3}$.

A key advantage 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. As expected, 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, in contrast to the F&S assumption, and the Spearman correlation coefficient confirms this ($rho = -0.03$, $p = 0.820$). We find that 23 out of our 61 subjects violate the F&S assumption that $\alpha_i \geq \beta_i$. They can be found to the left of the 45 degree line in the figure. Fourteen subjects do not exhibit inequality aversion in one of

[21] The reader will note that the top interval in Table 1 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".

the two directions. Nine subjects have $\alpha_i = 0$, five subjects $\beta_i = 0$, and one participant has $\alpha_i = \beta_i = 0$. Ten subjects are highly inequality averse in either direction. Eight subjects have $\alpha_i \geq 4.5$ and two subjects have $\beta_i = 1$.

5 Tests of the inequality aversion model

We now move on to test several hypotheses derived from the F&S model. Given the joint distribution of parameters of inequality aversion, we will sometimes analyze the results for a game in two steps. Whenever possible, 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 data we have and the analysis will be *as if* the data on the inequality parameters and those on the other decisions came from different and unrelated experiments. This is how previous test of the F&S model have proceeded. We will then go beyond that point by analyzing the individual-level data. Some of our hypotheses are only testable at the individual level though.

As for the statistical tools for our tests, we will use mainly non-parametric tests (chi-square test of proportions, Spearman correlation coefficient). Non-parametric test interpret the data in an ordinal fashion which we consider appropriate here. Occasionally, when a binary choice (e.g., *cooperate / defect*) is to be explained with the model's parameters (α, β), probit regressions may seem more appropriate even though this suggest a cardinal interpretation of the parameters. In such cases, we report the results of the probit regressions in footnotes.[22]

A crucial issue in our data analysis are the beliefs players hold about the inequality aversion (and resulting behavior) of the other players (although some hypotheses can be tested unconditionally, without specifying beliefs). In these cases, there are three possibilities. First, we can test what should happen when subjects have *correct beliefs* about the distribution of inequality parameters in the sample. Second, we often derive hypotheses for arbitrary *random beliefs* which are not correlated with players types. Third, we consider beliefs that are correlated with types, that is, we assume that a player i with a high α_i is more

[22] As we further discuss below, the probit analysis applies directly only to the linear inequality model of F&S, while the non-parametric tests that use only ordinal rankings apply equally to possible non-linear generalizations.

likely to believe player j to have a high α_j than a player i with a low α_i. In the social psychology literature such a 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) we will refer simply to a consensus effect. We study this alternative because we consider this the most plausible deviation from the hypothesis that all players hold the same beliefs.[23] Formal derivations of the hypotheses are presented in the Appendix.

5.1 Offers in the ultimatum game

We will start with the offers in the ultimatum game.

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

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 within-subject information we have available into account. In the data, we have 33 subjects with $\beta_i > 0.5$ and 26 subjects with $\beta_i < 0.5$.[24] In the UG, we observe 29 subjects who offer $z = 10$. The aggregate outcome of $z = 10$ 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 $z = 10$ observations is $(33 - 29)/33 = 12.1\%$ which seems a small enough deviation in order to accept the relevance of the F&S theory. More formally, we cannot reject that the share of subjects with $\beta_i > 0.5$ is the same as the share of subjects offering $z = 10$ ($\chi^2 = 0.544$, $d.f. = 1$, $p = 0.461$).

At the individual level, the data do not support the F&S model. Among the 33 subjects with $\beta_i > 0.5$, 18 chose $z = 10$, only slightly more than half of this group. A chi-square test on the $\beta_i > 0.5$

[23] See Engelmann and Strobel (2000) for evidence that subjects in an experiment with monetary incentives exhibit a clear consensus effect but no false consensus effect.

[24] There 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 model and so we discard them from the analysis. Note also that for no subject in our sample $\beta_i = 0.5$, so, we need only to distinguish $\beta_i \gtrless 0.5$.

observations cannot reject that choices are equiprobable ($\chi^2 = 0.273$, $d.f. = 1$, $p = 0.602$, one-sample test[25]).[26] Robustness checks with thresholds $\beta \in [0.3, 0.7]$ reveal that the insignificance result does not depend on the particular value of the $\beta = 0.5$ threshold. Therefore, we reject Hypothesis 1 (i).

We move on to the second part of the hypothesis. Among the 26 subjects with $\beta_i < 0.5$, 11 chose $z = 10$. This outcome, just as any other, is of course consistent with F&S but it seems remarkable that the share of subjects offering $z = 10$ here does not differ significantly to the one observed for the $\beta_i > 0.5$ subjects ($\chi^2 = 0.871$, $d.f. = 1$, $p = 0.351$, two-sample test). The behavior of the subjects with $\beta_i < 0.5$ is in principle consistent with Hypothesis 1 (ii) if subjects holds heterogenous beliefs. However, if we claim that the behavior of the $\beta < 0.5$ subsample is consistent, the aggregate outcome does not support F&S any more. The reason is that, if the $z = 10$ choices of $\beta < 0.5$ subjects are to be rational F&S choices, we should observe 44 $z = 10$ 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.

The previous argument highlights the role of beliefs for proposer behavior. It is possible to make predictions for general uncertain and for correct proposer beliefs.

Hypothesis 2 *Offers in the Ultimatum Game of subjects with* $\beta_i < 0.5$ *should be positively correlated with* β_i.

This hypothesis cannot be tested at the aggregate level. At the individual level, UG offers and βs should be positively correlated as long as beliefs (concerning the rejection probability) are not systematically (negatively) correlated with β. A Spearman test shows, however, no significant correlation ($rho = 0.187$,

[25] We 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.

[26] The reader will note that we accept the F&S model at the aggregate level when the chi-square test cannot reject that predicted and actual proportions are the same. At the individual level, we accept the F&S model only if there are significant differences in behavior between different parts of the sample (for example, $\beta \lesseqgtr 0.3$).

$p = 0.350$).[27] We restricted the test to the subjects with $\beta_i < 0.5$ because the other subjects should offer $z = 10$ 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 2 below). We conclude that the β data have no explanatory power regarding the UG offers.

We now consider the case where subjects hold correct beliefs.

Hypothesis 3 *If subjects know the true distribution of acceptance levels, all subjects should offer $z = 10$ in the Ultimatum Game.*

This hypothesis is clearly rejected from what was said above. On the one hand, the result may simply suggest that subjects' beliefs are wrong. On the other hand, this result may also indicate that the assumption of risk-neutral behavior underlying the F&S model is not appropriate here. Expected payoffs from offering $z \geq 5$ are relatively flat. Assuming $\beta_i = 0$ and given the responder behavior in our data, offering $z \in \{5, 8, 9\}$ yields an expected payoff of roughly 9.6, offering $z \in \{6, 7\}$ yields 9.0 on average, and offering the equal split always yields a payoff of 10. (Of course, expected utilities of players with $\beta_i > 0$ would be lower for offers $z < 10$.) Given these very similar expected payoffs, it seems that risk attitudes are a more plausible explanation for the heterogeneity of UG offers than inequality aversion. The flat expected payoffs also highlight the point made above that UG offers are not suitable to derive 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 $z < 10$ just as F&S assume. Even though F&S derivation of the β parameter could in principle be applied, Harrison's (1989) "flat maximum" critique would apply. It suggest that deviations from the actual maximum should be interpreted with caution when payoffs do not differ much around the maximum.

Hypothesis 4 *If subjects hold beliefs that are subject to a consensus effect, their Ultimatum Game offers should be positively correlated with α_i.*

[27] We report two-tailed p values throughout.

The hypothesis suggests a correlation between UG offers and UG rejection rates based on a consensus effect. It is consistent with F&S in the sense that subjects reject positive amounts in the UG but it is not an implication of that model because, given rejection behavior and a consensus effect, the hypothesis is based on simple payoff maximization. The intuition for this hypothesis is 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 as the Spearman rank correlation coefficient is significant ($rho = 0.398$, $p = 0.002$). We conclude that, in addition to differences in risk attitudes, differences in expectations about the behavior of responders appears to be a better predictor of the variation in UG offers than concerns for equal distribution of payoffs.

5.2 Contributions to the public good

Hypothesis 5 *(i) Subjects with $\beta_i < 0.3$ should choose $y_i = 0$ in the PG. (ii) Subjects with $\beta > 0.3$ may, depending on their beliefs, contribute any $y_i \in [0, 10]$ in the PG.*

We consider data at the aggregate level and take into account both merely positive contributions and contributions of at least half the endowment. There are 41 subjects with $\beta_i > 0.3$ and we observe 44 [36] subjects who contribute a positive amount [at least half the endowment]. 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. The formal test suggests that the proportion of zero [less than half] contributors is not significantly different from the one of $\beta_i < 0.3$ subjects ($\chi^2 = 3.49$, $d.f. = 1$, $p = 0.55$) [$\chi^2 = 0.88$, $d.f. = 1$, $p = 0.348$].

At the individual level, among the 20 subjects with $\beta_i < 0.3$, 13 [10] chose a positive contribution [at least half the 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), and neither are the proportions of contributors of less versus more than half the endowment ($\chi^2 = 0.00$, $d.f. = 1$, $p = 1.00$, one-sample). Among the other subjects, 31 out of 41 made a positive contribution [26 at least half the endowment]. This outcome is consistent with F&S. However, the difference to the subjects with $\beta_i < 0.3$ is not significant either 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 F&S theory.[28]

Hypothesis 6 *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 y_j 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_i. However, we cannot find support for the hypothesis. A Spearman correlation test indicates that the correlations have the right sign but neither the correlation between α_i and contributions ($rho = -0.177$, $p = 0.268$) nor between β_i and contributions ($rho = 0.104$, $p = 0.520$) are significant for the $\beta_i > 0.3$ subjects. The results do not change when we include also the subjects with $\beta_i < 0.3$, or when we include only subjects with a higher $\beta_i > \widetilde{\beta} \in [0.3, 0.6]$.[29]

Note that if beliefs are subject to a consensus effect, this would only strengthen the above hypothesis. Subjects with a higher α would expect others to contribute less and hence this would lower their utility maximizing contribution. In contrast, subjects with a higher β would expect others to contribute more which in turn increases their best response.

Hypothesis 7 *If subjects know the true distribution of PG contributions, subject i should choose the utility maximizing contribution given α_i and β_i.*

[28] For example, the share of subjects who contribute the full amount ($y_i = 10$) to the PG is virtually identical for the $\beta_i \gtrless 0.3$ subpopulations, 3 out of 20 and 8 out of of 41 respectively. Further, probit regressions on binary decisions to contribute to the public good (or contribute at least half the endowment) are not significant ($p = 0.481$, $p = 0.414$).

[29] Since 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.

The hypothesis says, somewhat tautologically, that all subjects will play their optimal contribution given their α_i and β_i if they are rational F&S players, and, since the true PG contributions are known, the utility maximizing individual contribution can be calculated. Put differently, it says that F&S is the correct model of behavior in this game. In order to analyze this hypothesis, we numerically derived each subject's optimal contribution.

At the aggregate level, the model's predictions are not impressive. F&S predict far too low contributions. For example, the predicted average contribution is 1.3 as opposed to the actual average of 4.7. Further, the F&S model predicts only 20 [12] positive contributors [contributions of at least half the endowment] whereas we observe 41 [36]. Finally, the model predicts no contribution above 7 but we have 16 subjects who contribute more than 7, eleven out of which contribute the full amount. A chi-square test comparing the proportion of contributors versus the non-contributors is highly significant ($\chi^2 = 18.93$, $d.f. = 1$, $p < 0.001$), so is the test on the proportion of subjects who contribute at least half the endowment ($\chi^2 = 13.66$, $d.f. = 1$, $p < 0.001$), suggesting that the model has no predictive power at the aggregate level.

If we take into account the within-subject data we have, we find a little explanatory power. The F&S model predicts 14 out of 17 zero contributors and 16 out of 44 contributors correctly. Whereas these differences in proportions are not significant ($\chi^2 = 2.00$, $d.f. = 1$, $p = 0.157$), the Spearman correlation of predicted and observed contributions is marginally significant ($rho = 0.22$, $p = 0.093$). This is an intriguing result as we saw above that α and β have no explanatory power (separately and jointly) for the PG contributions. 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. There is, however, a downside of this finding. All but two subjects predicted to make a positive contribution violate the $\alpha_i \geq \beta_i$ assumption of F&S. Put it another way, if we discard the $\alpha_i < \beta_i$ subjects from the analysis, no test whatsoever is significant regarding the above hypothesis.[30] We conclude that, even though it is consistent with F&S that subjects with a relatively high β contribute to the PG, these subjects should not exist according to one of the basic assumptions of the model. Moreover, the evidence we find is only marginally significant and therefore this result does not appear to be compellingly in favor of F&S.

[30] To be precise, no meaningfull test can be conducted as all but two subjects are predicted to free ride.

To sum up, we conclude that the F&S model has little explanatory power regarding the individual contributions to the PG. Limited predictive power of the model occurs in the case where subject have correct beliefs but it is driven by subjects who contradict a basic assumption of the model.

5.3 Behavior in the SPD

By backward induction, we start with second-mover behavior.

Hypothesis 8 *(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.*

Note that both parts of this hypothesis do not depend on beliefs. Therefore, behavior should unconditionally depend on the inequality parameters only, making the analysis simpler than in other cases.

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 a $(38 - 20)/20 = 90\%$. 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 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 parameters have some 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 ($\chi^2 = 3.97$, $d.f. = 1$, $p = 0.046$).[31] Even though this supports F&S, it should be emphasized that nearly half of our subjects $(22+4 = 26)$ violate part (i) of the hypothesis.

[31] A probit regression with the "cooperate" decisions as dependent variable and a dummy which is equal to 1 if and only if $\beta_i > 0.3$ yields the same result. A Spearman test correlating βs and "cooperate" decisions is significant ($rho = 0.293$, $p = 0.025$).

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 now turn to the first-movers. First-mover behavior depends on the beliefs of the subjects whether or not second movers will reciprocate cooperation. This was also the case in the previous sections but there we were still able to derive unconditional hypotheses for a subset of the participants (the $\beta_i > 0.5$ subjects in the UG and the $\beta_i < 0.3$ subjects in the PG). With the first movers in the SPD, this is not possible. If subject i believes with probability one that the second mover will reciprocate cooperation, i should cooperate regardless of the inequality 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.

If subjects hold non-degenerate beliefs, we obtain the following hypotheses.

Hypothesis 9 *First-mover cooperation decisions and α_i should be negatively correlated.*

This hypothesis can only be tested at the individual level. The Spearman correlation coefficient of individual i's first-mover "cooperate" decision and α_i is practically zero ($rho = -0.027$, $p = 0.840$). It appears that aversion against disadvantageous inequality does not have explanatory power regarding first-mover behavior.

Hypothesis 10 *If subjects know the true probability of second-mover cooperation, p, first-movers in the SPD should cooperate if and only if $\alpha_i < (7p - 3)/(10(1 - p))$.*

In the data, the actual value for p is $23/61 = 0.38$. Therefore, first-movers should cooperate if and only if $\alpha_i < -0.06$, that is, they should not cooperate. We observe 21 out of 61 subjects who cooperate as first movers, so, this is hypothesis not supported. This is not necessarily evidence against the F&S model as subjects may hold incorrect beliefs.

The β_i parameter does not seem to imply anything for first-mover behavior. The reason is that advantageous inequality for the first mover can only occur when the first mover defects and the second mover cooperates. This, however, should never occur according to the F&S model (and indeed this was nearly perfectly confirmed in our data). The β parameter may still imply a testable hypothesis though. A remarkable result in the SPD is that 45 out of 61 subjects make the same choice as first and second movers (after first-mover cooperation) and the correlation between these decisions is highly significant (see Table 2). This suggests that subjects' beliefs may be biased consistent with a consensus effect. That is, their first-mover decision is likely to be similar to how they decide as the second mover (a decision that does not depend on beliefs).

Hypothesis 11 *If subjects' beliefs are affected by a consensus effect, first-mover cooperation decisions and* β_i *should be positively correlated.*

We saw that the β distribution has some explanatory power regarding second movers and so one might expect the β parameter also to explain first movers via the consensus effect. However, the correlation is apparently too weak to carry over. The correlation coefficient we observe is not significant ($rho = 0.029$, $p = 0.828$). We conclude that first movers' expectations concerning second mover behavior appear to be substantially influenced by their own second-mover choice but this is not well captured by the F&S model.

5.4 Systematic patterns across games

We conclude the results section by reporting correlations across all decisions of the experiment. This is done, first, for the sake of completeness and, second, because the little explanatory power the inequality parameters may have suggested that individual behavior shows no consistent pattern at all across games. A reason for this could be that participants are confused by the multi-game setting and just 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 model could hardly be blamed for failing to predict individual decisions well.

Our data does, however, exhibit clear patterns. Table 2 presents the correlation coefficients across the decisions made in the experiment.[32] In each cell, the top entry is Spearman's correlation coefficient and the bottom entry is the two-tailed p value. We observe five significant correlations plus one that narrowly misses the 10% significance level.[33] Our conclusion from this is that behavior is not random or irrationally varied across decisions. At least some of the decisions in our experiment are correlated, and so the failure of the inequality model to predict individual decision across games must have different reasons.

This finding raises the question whether the observed correlations are predicted or in some other sense meaningful. In particular, we have to check whether they constitute evidence in favor or against the inequality aversion model of F&S.

The second move in the SPD (given first-mover cooperation) seems to have explanatory power for several other choices. This decision is positively correlated with UG offers, the first move in the SPD and contributions to the public good—in addition to the positive correlation with β already reported above. 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 against 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. However, it should be stressed that the correlation of first and second mover behavior in the SPD is a violation of the logic of inequality aversion. (We will discuss this point in detail in the next section.)

Next, we observe that UG offers and α (that is, UG acceptance levels) are positively and highly significantly correlated. This seems intuitive. A subject who rejects even moderately unequal offers will also tend to offer more herself. Similarly, a subject offering very little might also be expected to have a lower acceptance threshold. Andreoni et al. (2003) made the same comparison and found the same correlation

[32] We exclude the second move in the SPD given first-mover defection in Table 2. The reason is that virtually all subjects defect in this case, hence, this decision cannot reveal any insightful correlations.

[33] The correlation between the first move in the SPD and contributions to the public good has a p value of 0.11. A chi-square test with a binary decision in the PGG (contribute a positive amount or zero) and SPD first-mover decision is, however, highly significant ($\chi^2 = 5.36$, $d.f. = 1$, $p = 0.011$). The same result occurs when we analyze the binary decision whether subjects contribute at least half their endowment.

(we can compare our results to their "standard" treatment). The correlation between UG offers and α is not inconsistent with F&S but it does not confirm any prediction of the model either.[34]

Finally, we saw that the first move in the SPD and contributions to the public good can be considered as significantly correlated if one analyses a binary decision (whether or not to contribute a positive amount in the PG). In the F&S model, both decisions are to be associated with a low α parameter and therefore seem to confirm F&S. However, in our analysis above we saw that neither decision is explained by the α parameter. Moreover, we will see in the next section that subjects who contribute fully to the PG are significantly more likely to have a *high* (not low) α parameter. Therefore, the correlation of the first mover in the SPD and PG contributions are not evidence in favor of F&S.

To conclude, the correlations table shows that behavior is not erratic across the decisions of our experiment. Two of the correlations (in addition to the β correlation with SPD second mover) support the inequality aversion models but others are neutral or contradict the F&S model.

6 Discussion of the results

At the aggregate level, we found that the F&S model is consistent with UG offers and PG contributions. The model did not predict second mover behavior (given first-mover cooperation) in the SPD well. First mover decisions in the SPD turned out to depend crucially on beliefs, so, while the F&S does not explain these decisions well, we cannot exclude the possibility that the model is right but that subjects hold incorrect beliefs.

At the individual level, our main findings are (i) the α parameter has no explanatory power regarding the first move in the SPD and PG contributions; (ii) the β parameter has no explanatory power regarding UG offers and PG contributions; and (iii) the β parameter does have explanatory power regarding the second move in the SPD. We now discuss several possible explanations for our findings.

[34] Recall that F&S derive the β parameter from UG offers. In this case, the correlation of UG offers and α would support F&S's assumption that α and β are correlated. As mentioned above, we believe there are severe objections against this way of deriving the β. Furthermore, we do not find support that α and β are correlated.

We start with the first point. The following additional observations may be helpful in finding an explanation for the lack of predictive power of the α parameter regarding first moves in the SPD and contributions in the PG.

- In the PG, a look at the more extreme choices in the data suggests that contributions may actually be correlated in the direction opposite to the one the F&S model predicts. More precisely, subjects who contribute the full endowment are more likely to have a high α parameter (that is $\alpha_i > 2$) compared to the rest of the sample ($\chi^2 = 5.645$, $d.f. = 1$, $p = 0.018$).[35]

- Nearly half of our subjects behave consistently either "fairly" or "non-cooperatively" across the second move in the SPD (given first-mover cooperation), UG offers and UG responder behavior. There are 11 subjects who cooperate as second movers, offer the equal split in the UG and have an α_i higher than the median. We also observe 16 participants who appear to be selfish and non-cooperative. They defect as second movers, offer less than the equal split in the UG and have an α_i lower than the median. Crucially, when we compare only these "fair" versus the "non-cooperative" subjects, we find that the "fair" subjects are significantly more likely to cooperate as first movers in the SPD ($\chi^2 = 3.686$, $d.f. = 1$, $p = 0.055$). There is also a positive Spearman correlation to contributions to the PG (rho =0.34, $p = 0.083$). It may not seem surprising that "fair" subjects are more cooperative as first movers in the SPD and in the PG but this behavior is the opposite of what F&S predict as these subjects have a higher α_i.

- In the SPD, there is the above mentioned positive significant correlation of first and second mover behavior. This correlation is a violation of inequality aversion because inequality averse subjects (with a high α_i and β_i) should defect as first movers but cooperate as second movers. The opposite should hold for selfish subjects (with a low α_i and β_i). They should, if at all, cooperate as first movers and defect as second movers. As we have seen, our subjects do not conform to these predictions. As discussed above, in our data set, α and β are not correlated (in contrast to what F&S assume). Given

[35] Given that some subjects contribute a positive amount merely by "confusion" (Andreoni, 1995), such a test focussing on high contributions could be considered more appropriate.

this lack of correlation of the parameters, F&S predict no correlation between first and second move but then we do observe a significantly positive correlation.

Our conclusion from these observations is that inequality aversion does not capture what seems to be perceived as "fair" behavior here. Some subjects have a high rejection level as UG responders (high α_i) and nevertheless cooperate in the PG and as first movers in the SPD. Similarly, there seem to be subjects who are not particularly cooperative but they are not purely selfish either—otherwise, they should be more likely to cooperate as first movers. We think it is conceivable that subjects perceive some norm of "fairness" which does not coincide with inequality aversion in these cases. These subjects may conform to the norm or violate it but in either case their behavior is not consistent with F&S. This norm appears to encompass contributing to the public good, making and demanding fair offers in the UG, trusting and rewarding trust in the SPD. Some of these behaviors are in line with inequality aversion, others in clear contrast.

We now move on to finding (ii), the lack of explanatory power of the β parameter when it comes to UG offers and PG contributions. We believe that the reason why the β parameter does not explain UG offers and PG contributions is that some (if not most) subjects behave differently in strategic games as opposed to simple distribution games. This argument is not novel. Various authors have suggested that UG offers are not driven by inequality aversion or altruism but that players behave strategically (e.g., Forsythe et al., 1994). Camerer (2003, p.56)[36] 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.

[36] For a similar argument, see Charness and Rabin (2002). Fehr, Naef and Schmidt (2006) also make the argument that subjects behave differently in distribution games and strategic games.

Our results regarding UG offers are consistent with this interpretation for the following reason. Subjects with $\beta_i > 0.5$ have a switching point smaller than 10 in the MDG. This decision to switch below 10 costs these subjects more money than the other player receives. In other words, they are willing to pay a price higher than one for each unit that the recipient receives. In the UG, every monetary unit the responder receives costs the proposer exactly one. Hence, one would expect the subjects with a switching point below 10 to offer $z = 10$ in the UG. As seen above, this is often not the case, and there is even no correlation between β_i and the offer in the UG. What seems to be happening is that the (15 out of 33) subjects who violate Hypothesis 1(i) are more generous in the MDG compared to the UG where they face the risk of rejection. We believe this is the key for understanding our results regarding the lack of explanatory power of the β parameter.

Another observation consistent with this argument is the finding that α and β are not correlated but UG offers and rejection levels are. This also suggests that the UG offers are not driven by inequality aversion. We suspect that they are primarily driven by the expectation concerning the rejection probability and that players with a higher rejection threshold also expect a higher probability of offers being rejected, which is consistent with a consensus effect.

Based on the argument that strategic and distribution games differ, one could argue that the inequality parameter should be estimated from a strategic game. If, despite the problems with this approach discussed above, we tried to identify the β parameter from the UG offers (as F&S do), support for F&S would not improve. In Table 2, we find a positive correlation of UG offers with SPD second mover cooperation. Since second mover cooperation is also correlated with out measure of β, this approach does hence not yield additional support for F&S.

Finally we come to point (iii), why does the F&S model have some predictive power in the second move of the SPD then? Here the player makes a final decision over outcomes, so the correlation with the MDG is no coincidence. In Camerer's interpretation, the deciding player is responsible for the other player—just as in the MDG. Accordingly, the β parameter has predictive power. More generally, what we observe in our data is that a large share of the observed behavior is correlated across games. What does not seem to be correlated is behavior when subjects make final decisions over payoff distributions (UG responder, or α;

MDG, or β) and behavior in strategic situations (SPD first move, PG, UG proposer).[37] We conclude that it appears that other considerations can dominate purely distributional concerns in strategic games and, importantly, that these considerations are not correlated in a systematic way with distributional concerns.

It is by now accepted that inequality aversion cannot explain all games.[38] Several papers have shown that, among other motives, intentions (see e.g. Falk, Fehr, and Fischbacher, 2003) and efficiency concerns (see e.g. Charness and Rabin, 2002, Engelmann and Strobel, 2004) play an important role in explaining some experimental results. In our experiments, we focussed on games that could well be rationalized by inequality aversion, and have indeed even inspired the inequality model. Nevertheless, we have to consider whether other motives might partly drive our results.

Intentions are clearly irrelevant in the MDG. In contrast, rejections in the UG could be driven by negative reciprocity (and the evidence in Falk, Fehr, and Fischbacher, 2003, for mini-ultimatum games suggests that they are to a large extent). As a result, our estimates of α might be biased. Nevertheless, if the inequality model is supposed to work as an "as if" model in a large class of games by capturing both literal inequality aversion as well as negative reciprocity, our estimates of individual α's should still have predictive power. This is true unless negative reciprocity and inequality aversion are unrelated within subjects. We get back to this issue below. Similarly, second mover SPD behavior can be influenced by positive (given first mover cooperation) and negative (given first mover defection) reciprocity.

Reciprocity might also explain why second mover cooperation correlates both with UG offers and β, while UG offers and β are uncorrelated. As discussed above, second movers make final decisions over payoffs so a correlation of cooperation with distributional concerns as measured by β does not come as a surprise. On the other hand, second mover cooperation is clearly consistent with (positive) reciprocity. In

[37] The exception is the final decision made by SPD second movers.

[38] Several commentators on our paper encouraged us to check whether Quantal Response Equilibrium (McKelvey and Palfrey, 1998) can explain our data. While we are sympathetic to this idea, we believe it is virtually impossible to conduct such an analysis. For a start, there are seven different decisions to take into account even if we count the 21 UG and MDG decisions as one decision. What is more, virtually 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 best response against a probability distribution over the types of players whom i faces. Now for each type of player in this distribution, there is another probability distribution across the various choices this type of player might take. Conduction a QRE analysis, therefore, seems a formidable task.

the UG (negative) reciprocity implies higher rejection rates so a proposer who considers reciprocity to be an important motive for the responder will make higher offers. Consequently, a subject who is reciprocal and expects other to be so as well would cooperate as second mover (after first-mover cooperation), but would also make a relatively high offer in UG. This would imply that second-mover cooperation and UG offers are correlated for a different reason than second-mover cooperation and MDG giving (and the underlying motives themselves appear to be uncorrelated).

Concerns for efficiency may play a role in all four games. In the UG, rejecting an offer not only decreases the inequality between proposer and responder, it also burns the entire £20 pie. Therefore, a subject who is concerned with overall efficiency may be less inclined to reject an offer. This implies that our measure of α would be biased downwards if subjects care for efficiency. In the SPD and in the PG, cooperation is not individually profitable but even unilateral cooperation increases the sum of payoffs of the two players. Therefore, participants with a preference for efficiency should, all else equal, cooperate more. In these three games, efficiency concerns are partly in conflict with individual profit maximization and with inequality aversion. This does not, however, affect our hypotheses with respect to the correlations. Moreover, efficiency concerns have been invoked in distribution experiments where the inequality model failed to capture choices that increase the payoff of players that already are better off. In principle, this could be captured by allowing for $\alpha < 0$, so that a generalized inequality model[39] could capture this motivation as well. In the MDG, efficiency concerns might also play a role. When a player chooses the egalitarian outcome below a level of £10, efficiency is reduced. By contrast, choosing the egalitarian outcome at any point above £10 increases efficiency. If players have efficiency concerns, their choice should be biased upwards whenever their unbiased switching point is below £10 and biased downwards if their "true" switching point is above £10. Note, however, that as long as efficiency concerns are not in some systematic way correlated with inequality aversion, the resulting bias in the estimate of β would not affect the expected correlations with other behavior.

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,

[39] Such a model would be a linear version of the altruism model of Cox and Sadiraj (2005).

our main conclusions are based on the absence of a correlation between the inequality parameters that we estimated based on behavior in the UG and DG 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 correlation between the switching point and the measure of inequality aversion is 1. Furthermore, stronger inequality aversion has again the same implications for the decisions in the other games, e.g., stronger aversion towards disadvantageous inequality implies lower contribution levels in the PG. Now the absence of a correlation between our estimates of the inequality parameters in the linear model and the behavior in the other decision nodes means nothing but that the switching points in the MDG and UG are not correlated with the other behavior. Since the latter are perfectly correlated with any non-linear measure of inequality aversion, this means that such measures would also not be significantly correlated with the behavior in the other decision nodes. 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.[40]

7 Conclusions

In this paper we assess the predictive power of 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 of the experiments. We use the responder data from the ultimatum game in order to estimate a parameter of aversion against disadvantageous inequality, and we take data from a modified dictator game to estimate a parameter of aversion against advantageous inequality. We then use this joint distribution to test several hypotheses about individual behavior in the other games.

[40] This is not to say that a generalization would not improve the predictive power of F&S in other instances, in particular in multi-player games. Furthermore, it might improve the success of F&S in our experiment at the aggregate level, but it does not influence our crucial conclusions for the individual level data.

Our results show that the inequality aversion model has some predictive power at the aggregate level but fails almost entirely at the individual level. At the aggregate level, we found that the F&S model is consistent with UG offers and PG contributions but the model did not predict second-mover behavior (given first-mover cooperation) in the SPD well. At the individual level, we found that the α parameter cannot explain the first move in the SPD and PG contributions, and that the β parameter is inconsistent with UG offers and PG contributions. However, we also found that the β parameter is consistent with second-mover behavior in the SPD.

Our first conclusion is that aggregate support of a theory, if remarkable, should not be equated to individual validity of the theory. This seem particularly relevant for behavioral models of other-regarding preferences. A second issue is that we found two cases (contributions to the public good and first-mover behavior in the sequential prisoner's dilemma) where inequality aversion predicts the inequality averse subjects to defect or free ride—which is contradicted by our data. Third, our results and discussion suggest that an inequality model calibrated on distributional decisions has little predictive power in strategic situations.

It appears to us that the success of the inequality model at an aggregate level is largely based on its ability to qualitatively capture different important motives in different games, including altruism in the dictator game, (negative) reciprocity in the ultimatum game, (positive) reciprocity in the sequential prisoner's dilemma game. To some extent, this is supported by our data. The low predictive power of the model at an individual level then seems to be driven by the low correlation of these motives within subjects. Thus it appears to be both the strength and the weakness of the inequality model that it can capture different motives in one functional form. On the one hand, this permits to rationalize several apparently disparate results in one simple model. On the other hand, an individual's behavior is not well captured by this same model, since different motives drive behavior in different situations and these seem to have little correlation within subjects. The inequality model can hence serve as a relatively elegant "as if" model in several situations one at a time, but it does not appear to accurately and consistently reflect the motives of individuals.

Concluding, we stress that while we have applied our within-subject test to the inequality aversion model by F&S, the main conclusion we draw is that any theory of other-regarding preferences should pass such a test in order to be considered a reliable explanation of experimental behavior. The main reasons we have tested F&S here were practical considerations and the success and attention it has achieved in the past.

References

[1] 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.

[2] Andreoni, J. (1995). "Cooperation in Public Goods Experiments: Kindness or Confusion?" American Economic Review 85, 891-904.

[3] Berg, Joyce, John Dickhaut and Kevin McCabe (1995). "Trust, Reciprocity and Social History." Games and Economic Behavior 10, 122-42.

[4] Bolton, Gary E. (1991). "A Comparative Model of Bargaining: Theory and Evidence." American Economic Review 81, 1096-136.

[5] Bolton G. E. and Zwick R. (1995). "Anonymity versus Punishment in Ultimatum Bargaining." Games and Economic Behavior 10, 95- 121.

[6] Bolton G. E. and Ockenfels A. (2000). "ERC: A theory of Equity, Reciprocity and Competition." American Economic Review 90, 166-193.

[7] Camerer C. (2003). "Behavioral Game Theory: Experiments in Strategic Interaction (Roundtable Series in Behavioral Economics)." Princeton University Press.

[8] Charness, G. and Rabin, M. (2002). "Understanding Social Preferences with simple tests." Quarterly Journal of Economics 117(3), 817-869.

[9] Clark K., and Sefton, M. (2001). "The sequential prisoner's dilemma: evidence on reciprocation." The Economic Journal 111, 51- 68.

[10] Cox J. C., Friedman, D., Gjerstad, S. (2005). "A tractable model of reciprocity and fairness." Mimeo.

[11] Dufwenberg, M and Kirchsteiger, G. (2004) "A theory of sequential reciprocity." Games & Economic Behavior 47, 268-98.

[12] Engelmann, D. and Strobel, M. (200) "The False Consensus Effect Disappears if Representative Information and Monetary Incentives Are Given." Experimental Economics 3(3), 241-60.

[13] Engelmann, D. and Strobel, M. (2004) "Inequality Aversion, Efficiency and Maximim Preferences in Simple Distribution Experiments." American Economic Review 94(4), 857-69.

[14] Engelmann, D. and Strobel, M. (2006) "Subject Pool Effects in the Importance of Concerns for Equality and Efficiency - A Reply to Fehr, Naef and Schmidt." American Economic Review, forthcoming.

[15] Falk, A., Fehr, E., and Fischbacher, U. (2003) "On the Nature of Fair Behavior." Economic Inquiry 41(1), 20-26.

[16] Falk, A. and Fischbacher, U. (2006) "A Theory of Reciprocity." Games and Economic Behavior 54(2), 293-316.

[17] Fehr, E. Naef, M., and Schmidt, K.M. (2006). "The Role of Equality and Efficiency in Social Preferences." American Economic Review, forthcoming.

[18] Fehr, E. and Schmidt, K. M. (1999). "A theory of Fairness, Competition and Cooperation." Quarterly Journal of Economics 114(3), 817- 868.

[19] 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.

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

[21] Güth, W.Schmittberger, R. Schwarze B. (1982). "An experimental analysis of ultimatum bargaining." Journal of Economic Behavior and Organization 3, 367- 388.

[22] Harrison, G. (1989). "Theory and Misbehavior of First-Price Auctions." American Economic Review 79, 749-762.

[23] Holt, C.A., Goeree, J. and Laury, S. (2002). "Private Costs and Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior." Journal of Public Economics 83(2), 257-278.

[24] Kahneman, D., Knetsch, J., and Thaler, R. (1986). "Fairness and the assumptions of economics." Journal of Business 59, S285-S300.

[25] Ledyard, J.O. (1995). "Public Goods: A survey of experimental research." The Handbook of Experimental Economics, edited by John H Kagel and Alvin E. Roth. Princeton University Press.

[26] Levine, D. (1998). "Modelling Altruism and Spitefulness in Experiments." Review of Economic Dynamics 1, 593-622.

[27] McKelvey, R. D. and Palfrey, T.R. (1998) "Quantal Response Equilibria for Extensive Form Games." Experimental Economics 1, 9-41.

[28] 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.

[29] Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics." American Economics Review 83(5), 1281- 1302.

[30] Roth A.E. (1995). "Bargaining Experiments." The Handbook of Experimental Economics, edited by John H Kagel and Alvin E. Roth. Princeton University Press.

[31] Shaked, A. (2005). "The Rhetoric of Inequity Aversion." University of Bonn working paper.

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* $z_i = 10$ *in the Ultimatum Game. (ii) Subjects with* $\beta_i < 0.5$ *may, depending on their beliefs, offer either* $z_i = 10$ *or* $z_i < 10$ *in the Ultimatum Game.*

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

Hypothesis 2 *Offers in the Ultimatum Game of subjects with* $\beta_i < 0.5$ *should be positively correlated with* β_i.

Proof. Player i believes that an offer z is accepted with probability $d_i(z)$. Following F&S, we have $d_i(z) \leq d_i(z + 1)$ and $d_i(10) = 1$. Offering z yields an expected utility to the proposer of

$$d_i(z)\,(20 - z - \beta_i(20 - 2z))\,.$$

A higher β_i implies that $(20 - z - \beta_i(20 - 2z))$ decreases less in z, which, all else equal, implies a (weakly) higher optimal offer. ■

Hypothesis 3 *If subjects know the true distribution of acceptance levels, all subjects should offers* $z = 10$ *in the Ultimatum Game.*

Proof. This follows from our responder data—see the expected proposer payoffs reported in the main text. ■

Hypothesis 4 *If proposers hold beliefs that are subject to a consensus effect, their offers should be positively correlated with* α_i.

Proof. All other things equal, if subjects exhibit a consensus effect, a higher α_i implies a (weakly) higher expected rejection probability for any possible offer. This implies that the payoff maximizing offer will be higher. ■

In the following proofs regarding the PG, we use notation for individual contributions, y_i, not employed in the main text.

Hypothesis 5 *(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 $\overline{y} \in [0, 10]$ so that the payoff for player i is $10 - y_i + 0.7(y_i + \overline{y}) = 10 + 0.7\overline{y} - 0.3y_i$ and the payoff of player j is $10 + 0.7y_i - 0.3\overline{y}$. If player i also contributes $\overline{y}$, he gets a utility of $10 + 0.4\overline{y}$. If player i contributes $y_i < \overline{y}$, this yields a utility of $10+0.3(\overline{y}-y_i)+0.4\overline{y}-\beta_i(\overline{y}-y_i)$ which is larger than $10+0.4\overline{y}$ if and only if $\beta < 0.3$. If player i contributes $y_i > \overline{y}$, this yields a utility of $10 - 0.3(y_i - \overline{y}) + 0.4\overline{y} - \alpha_i(y_i - \overline{y}) < 10 + 0.4\overline{y}$. Hence, player i will never contribute more than $\overline{y}$, will contribute $\overline{y} \in [0, 10]$ if $\beta_i > 0.3$, and will contribute $y_i = 0$ if $\beta_i < 0.3$. ■

Hypothesis 6 *Contributions to the PG of subjects with* $\beta_i > 0.3$ *should be negatively correlated with* α_i *and positively correlated with* β_i*.*

Proof. Suppose player i believes contribution level y_j will be chosen with probability $d_i(y_j) \in [0,1]$, where $\sum_{y_j=0}^{10} d_i(y_j) = 1$. Using the expressions derived above, i's expected utility from contributing y_i is

$$\begin{aligned}
&\sum_{y_j=0}^{y_i-1} d_i(y_j)[10 - 0.3(y_i - y_j) + 0.4y_j - \alpha_i(y_i - y_j)] + d_i(y_i)[10 + 0.4y_i] + \\
&\sum_{y_j=y_i+1}^{10} d_i(y_j)[10 + 0.3(y_j - y_i) + 0.4y_j - \beta_i(y_j - y_i)] \\
= \; &\sum_{y_j=0}^{10} d_i(y_j)(10 + 0.4y_j) - \sum_{y_j=0}^{y_i-1} d_i(y_j)(0.3 + \alpha_i)(y_i - y_j) + \sum_{y_j=y_i+1}^{10} d_i(y_j)(0.3 - \beta_i)(y_j - y_i).
\end{aligned}$$

From this expression, the higher β_i, the more do own contributions y_i increase the expected utility (provided $\beta_i > 0.3$), and vive versa for α_i. ■

Hypothesis 7 *If subjects know the true distribution of PG contributions, subject* i *should contribute the utility maximizing given* α_i *and* β_i*.*

Proof. Suppose the actual relative frequency of contribution level y_j is $f(y_j)$ and N is the number of players in the experiment. Using the terms from the proof of Hypothesis 6, we obtain the expected utility from contributing y_i as

$$\begin{aligned}
&\sum_{y_j=0}^{y_i-1} [10 - 0.3(y_i - y_j) + 0.4y_j - \alpha_i(y_i - y_j)]f(y_j) + [10 + 0.4y_i](f(y_j) - 1/(N-1)) + \\
&\sum_{y_j=y_i+1}^{10} [10 + 0.3(y_j - y_i) + 0.4y_j - \beta_i(y_j - y_i)]f(y_j).
\end{aligned}$$

The optimal contribution is the one yielding the highest expected utility. ■

Hypothesis 8 *(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 $14 < 17 - \beta_i(17-7) \Longleftrightarrow \beta_i < 0.3$. (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 $\alpha_i > 0$. ■

Hypothesis 9 *First-mover cooperation decisions and* α_i *should be negatively correlated.*

Proof. Suppose first-mover i believes that the probability of the second mover to cooperate is p_i. As shown in the proof of hypothesis 10 below, cooperating yields an expected payoff higher than defecting if and only if $\alpha_i < (7p_i - 3) \,/\, (10(1-p_i))$. Therefore, first-mover cooperation decisions and α_i are negatively correlated (unless beliefs and types are correlated). ■

Hypothesis 10 *If subjects know the true probability of second-mover cooperation, p, first-movers in the SPD should cooperate if and only if* $\alpha_i < (7p-3)/(10(1-p))$.

Proof. If the first mover defects, the second mover will also defect (regardless of α_j and β_j) and both players get $U_i(10,10) = 10$. The first mover's belief for the second mover to cooperate is 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

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

■

Hypothesis 11 *If subjects' beliefs are affected by a consensus effect, first-mover cooperation decisions and* β_i *should be positively correlated.*

Proof. If subject exhibit a consensus effect, beliefs about the probability of the second mover to cooperate (p_i) are a function of β_i such that $\partial p_i(\beta_i)/\partial\beta_i > 0$. Cooperating yields an expected payoff higher than defecting if and only if $\alpha_i < [7p_i(\beta_i) - 3]/[10 - 10p_i(\beta_i)]$. Therefore, a higher β_i implies that i is more likely to cooperate. ■

Characterization of the MDG

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

$$x_i - \beta_i(x_i - x_j) = x_i' - \beta_i(x_i' - x_j'). \tag{4}$$

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 should avoid the trivial solution where $(x_i, x_j) = (x_i', x_j')$. Third, we need $sign(x_i - x_i') = sign(x_i - x_j - (x_i' - x_j'))$ 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

$$x_i = x_i' - \beta_i(x_i' - x_j') \tag{5}$$

or

$$\beta_i = \frac{x_i' - x_i}{x_i' - x_j'} \tag{6}$$

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 $x_i \in \{0, 1, 2...20\}$). 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 elilict the β_i parameter. In particular, it also allows us to detect whether there are, in violation to the F&S assumptions, 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.[41] 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 $k/(k + x_i - x_j') < 1$.[42] 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.

[41] 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.

[42] 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.

Tables

α			β		
	F&S	data		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 1. 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 (0.82)	0.40 (0.00)***	0.07 (0.60)	-0.03 (0.84)	0.16 (0.22)
β		—	0.13 (0.31)	0.13 (0.30)	0.03 (0.83)	0.29 (0.02)**
UG offer			—	0.19 (0.15)	0.13 (0.31)	0.43 (0.00)***
PG				—	0.20 (0.11)	0.35 (0.01)***
SPD 1^{st}					—	0.34 (0.00)***
SPD 2^{nd}						—

Table 2. Spearman rank correlations between decisions (two-tailed p value in parenthesis)

FIGURE 1: The joint a-ß distribution