Inequity Aversion Revisited

Yang Yang*
Sander Onderstal**
Arthur Schram,*****

***Sun yat-Sen University, Guangzhou, China**
****University of Amsterdam and Tinbergen Institute, The Netherlands**
*****Robert Schumann Center for Advanced Studies, European University Institute, Florence**

November 11, 2015

ABSTRACT:
Using a laboratory experiment, we study the predictive power of the Fehr-Schmidt (1999) model of inequity aversion and its robustness to reciprocity and stakes. We find stronger evidence for the model's predictive power at the individual level than what the existing literature suggests. This finding is robust to stakes. However, the model's predictive power is highly reduced if subjects can reciprocate others' actions. This suggests that parameter estimates obtained in an environment that allows for reciprocal responses yield a bias in the parameter estimates. In particular, previous estimates (especially of the disutility of disadvantageous inequity aversion) may overestimate the importance of inequity aversion.

KEYWORDS: Inequity aversion, stakes, reciprocity, robustness

***JEL* CODES:** C52, C91, D63

Please address correspondence to: Yang Yang, Lingnan College, Sun Yat-Sen University, No. 135, Xingang Xi Road 510275, Guangzhou, China

Email:
Yang: yangyuu@gmail.com
Onderstal: Onderstal@uva.nl
Schram: Schram@uva.nl .

1. Introduction

Why are you reading yet another paper on inequity aversion (IA)? One would think that the plethora of papers that have studied this phenomenon since the seminal work by Fehr and Schmidt (1999) would have said it all.[1] In our view, this is not the case. We believe that at least two types of questions have remained underexplored in this extensive literature: First, how robust are the estimated parameters of the Fehr and Schmidt's (1999) IA model ("the IA model" henceforth) at the individual level and, second, how stable is the accuracy of the predictions these parameters yield? In particular, does it matter for the estimations whether or not people have an opportunity to reciprocate others' actions in the environment used for the calibration? And, how sensitive is the parameter distribution to variations in stakes? Moreover, how does the predictive power of the IA model at the individual level vary across environments? More specifically, does a model calibrated in one environment predict well in an alternative task where the situation is different with respect to reciprocity opportunities or stakes?

Such questions pertain directly to the empirical applicability of the IA model. Given that this model has been used to explain people's behavior in many experimental and real-life environments and is even seen as a preferred first approach for studying behavior in many areas (Camerer 2003:472), a systematic analysis of its strengths and weaknesses seems of utmost importance. This kind of 'scientific' approach to critically assessing theories was recently advocated by Binmore and Shaked (2010) for economics in general and for the IA model in particular. Their

[1] Alternative prominent IA models include Bolton and Ockenfels (2000) and Charness and Rabin (2002). See Cooper and Kagel (2012) for a review of the literature on inequity aversion.

main focus is on the methodology used by Fehr and Schmidt, however.[2] Though we agree that a critical assessment of methods is important, it is also necessary to test a model's predictive power, especially when it is so widely used: Binmore and Shaked report that as of 2010 Google Scholar listed 2390 works citing Fehr and Schmidt; by November 2015 this had increased to almost 8,000. Only a few of these (to be reviewed below) investigate the model's predictive power at the individual level or systematically study its robustness to stakes and to the opportunity to reciprocate.

This paper addresses both types of questions in a laboratory experiment. More specifically, we use a three-step design to test the IA model. The first step is to estimate each subject's IA preferences by using a set of novel choice menus. In the second step, we check the robustness of these estimations to scaling (i.e., we test the linearity assumption underlying the model) by varying the stakes in the menus. Note that the validity of the linearity assumption will be an important issue in any extrapolation of the model from the laboratory to the field. In the third step, we let subjects play the 'production game'. We developed this game to create various desirable properties (to be discussed below) that will allow us to straightforwardly compare subjects' decisions to the theoretical prediction derived from the estimates of their individual IA levels as obtained in the first two steps. This allows us to test the predictive power of the IA model and to check the robustness of the predictive power to reciprocity opportunities and stakes.

Of course, there have been other attempts to test the IA model. In fact, economists have conducted various laboratory experiments to test it and, typically, to compare it to other social preference models (e.g., Charness and Rabin 2002, Engelmann and

[2] Importantly, we do not intend to 'take sides' in the specific criticisms that Binmore and Shaked put forward with respect to the Fehr-Schmidt model. See Fehr and Schmidt (2010) for a concise reply to these critiques.

Strobel 2004, and Brandts et al. 2014). Typically, such studies develop a series of simple games for which various models (including the IA model) offer distinct predictions. Subjects' choices in these games subsequently provide evidence in favor or against specific models. In this way, it has been argued that efficiency concerns (Charness and Rabin 2002) and maximin preferences (Engelmann and Strobel 2004) are better predictors of individual choice than the IA model (though the latter result has been disputed by Fehr et al. 2006). An advantage of this method (as pointed out by an anonymous referee) is that it allows one to test various models in one setting. Notwithstanding the elegance and usefulness of such horse races between models, they do have drawbacks, however. Most importantly, not the models themselves are tested, but their comparative predictions. Our approach is complimentary in that we will directly test the model's premises with respect to preferences and its predictions, independently of other models.

Originally, many of the experimental tests (including the supporting evidence in Fehr and Schmidt 1999) were implemented at an aggregate level, by checking the consistency of the distribution of subjects' behavior across different games with the distribution predicted by the model. This method has an important drawback. Even if the model passes the test at the aggregate level, this is not informative about its predictive power for each individual. More recent studies, including Engelmann and Strobel (2004) and Blanco et al. (2011), have recognized this problem and test predictions at the individual level. Such tests can directly check the within-subject consistency for each individual. Blanco et al.'s (2011) approach is probably closest to ours.[3] As our paper, their work is an individual-level study of the IA model that tests the internal consistency of the IA model across different games. Their method to

[3] Dannenberg et al. (2007) present an application of the Blanco et al. method.

measure the individual guilt level (i.e., the disutility derived from having a higher payoff than others) is based on a modified dictator game that has a structure similar to the menu we use to measure guilt.[4]

Our approach has several advantages over previous individual-level tests of the IA model. First, in our production game, a player's behavior predicted by the IA model is a continuous function of her IA levels and each player's predicted behavior depends on only one type of IA (either advantageous or disadvantageous). In the games traditionally used to test the model (e.g., the prisoners' dilemma or the ultimatum game), the model's prediction is a binary function, e.g. cooperate/defect or accept/reject, of the player's IA levels. The production game we introduce facilitates a sharper test of the model because it avoids such bang-bang predictions.

Second, in our production game, subjects' risk attitudes and beliefs on the distribution of IA preferences in the population are irrelevant because under the IA model players have dominant strategies. This is in contrast to most of the games traditionally used for the analysis of IA preferences (e.g., the public good game, the prisoners' dilemma, and the ultimatum game; see, e.g., Fehr and Schmidt 1999 and Blanco et al. 2011). These traditional games are characterized by strategic uncertainty; hence, a subject's decisions may be affected by her risk attitude and beliefs, which could yield biased estimates of her IA preferences.[5] Our experiment allows for a more

---

[4] Like Goeree and Holt (2000) and Blanco et al. (2011), we use the term 'guilt' to indicate disutility derived from earning more than others. This is not to be confused with the same term referring to disutility derived from failing to meet others' expectations (e.g., Charness and Dufwenberg 2006).

[5] For instance, a risk averse proposer in an ultimatum game may choose a high offer because she fears rejection by a receiver with a high disutility of disadvantageous IA. If this is not taken into account when analyzing her choices (e.g., Fehr and Schmidt 1999; Blanco et al. 2011), her high offer may yield an erroneously high estimate of disutility of advantageous IA. Recently, some studies have attempted to investigate inequity aversion in the absence of strategic uncertainty. For instance, Teyssier (2012) and Blanco et al. (2014) use a sequential prisoner's dilemma game and a public good game, respectively, to study first and second mover decisions. Nevertheless, first movers still face the uncertainty of what type they are faced with. Moreover, reciprocity may play a role in the second mover's decisions.

accurate test of the IA model as risk attitudes and beliefs regarding other players' IA preferences play no role.

Third, by comparing two versions of the production game that only differ in the opportunity to respond to the other worker's decision, we can isolate the role of 'explicit reciprocity' (Charness and Rabin 2002) from IA preferences in a way not previously done. Previous studies have established the importance of taking reciprocal responses to fairness intentions into account above and beyond straightforward preferences for fairness (e.g., Falk et al. 2008). For instance, ultimatum game experiments have revealed that the responder's choice may depend heavily on the actions available to the proposer, a finding that is not captured by the IA model.[6] In other words, the distinction between inequity aversion *per se* and (preferences for) reciprocity is important. This means that a proper test of the model's predictive power should carefully control for the possibility of reciprocity.[7] Therefore, instead of deriving subjects' levels of envy (i.e., disutility of disadvantageous inequality) indirectly from the ultimatum game as in Blanco et al. (2011), we use a set of simple menus to directly measure them.

Finally, we are the first to systematically test the IA model's robustness to variations in stakes in a setting where players cannot reciprocate others' choices.[8]

Our findings show, first, that the parameter estimates of the inequity aversion model are– to a large extent– robust to variations in the stakes.[9] Second, the IA

[6] See, e.g., Brandts and Solà (2001) and Falk et al. (2003).
[7] In settings where players cannot reciprocate others' actions, subjects could still reciprocate based on beliefs about an opponent's action in the sense of Rabin (1993). Throughout the paper, we will speak about "explicit reciprocity" or "reciprocity opportunities" to indicate environments where players can condition their choices on the *actual actions* by others.
[8] See Carpenter et al. (2005), List and Cherry (2008), Anderson et al. (2011), and Fehr et al. (2014) for examples of robustness tests with respect to stakes in environments where such explicit reciprocity is possible.

model's predictive power is also robust to variations in stakes. Third, the IA model predicts individual subjects' behavior in the production game quite well in a reciprocity-free setting. Adding the opportunity to reciprocate to the game strongly reduces the model's predictive power, however. Finally, our estimate of the distribution of the guilt parameter is roughly similar to those reported in previous studies, including Fehr and Schmidt (1999), but we find much lower levels of envy. However, when deriving estimates of envy from individuals' choices in the version of the production game that allows for reciprocity, our estimates of subjects' envy levels are higher and correspond better to the previous estimates.

Our experimental observations thus suggest that concerns for intentions and the possibility to reciprocate others' actions can to a large extent explain the discrepancy between our results and others reported in the literature. Note that Fehr and Schmidt (1999) already point out that their parameters may be interpreted as a combination of concerns for both inequity and intentions; e.g., a rejection of a positive but low ultimatum offer may be interpreted as aversion to the resulting inequity or a reciprocal response to the perceived bad intention of the proposer. We believe that our methods are the first to facilitate a clear separation of the intention-based concern from the preference for equality.

The remainder of the paper is organized as follows. In Section 2 we will introduce the models, including the details and predictions for the production game. We will explain the experimental design in Section 3 and the results in Section 4. Finally, in Section 5, we offer a discussion of the results and a conclusion of the paper.

---

[9] In online appendix D, we show that our results are also not affected by allowing subjects to have efficiency concerns. The online appendices are available at […].

2. Theory

We will derive hypotheses using the classic IA model (Fehr and Schmidt, 1999). In this section, we recall the model and introduce the production game that we will use to measure its predictive power.

Consider a population of n+1 individuals. In the IA model, an individual derives positive utility from own earnings and (dis)utility from inequality. More specifically individual i's utility is given by:

Inequity aversion model:

$$U_i(x, y_1, y_2, \ldots, y_n) = x - \alpha_i \frac{\sum_j \max\{y_j - x, 0\}}{n} - \beta_i \frac{\sum_j \max\{x - y_j, 0\}}{n} \qquad (1)$$

where x is player i's material payoff, and y_j, $j = 1, \ldots, n,$ denote the other n players' payoffs. $\alpha_i \geq 0$ is player i's envy parameter (Engelmann and Strobel 2004) measuring the marginal disutility of disadvantageous inequality. β_i measures the marginal disutility (guilt) related to advantageous inequality, with $\beta_i \in [0, \alpha_i] \cap [0,1)$.[10]

Next, we introduce the production game that we will use to test the predictive power of the IA model. This game is played by two players, Worker A and Worker B. At the start of the game, each receives a basic salary (s_i, i = A, B). This represents an initial income from which each worker can pay her effort costs. The effort can yield a bonus as described below.

Each worker is in charge of a department's production (departments are also denoted by A and B). The production of each department will be equally distributed (as a 'bonus') between the two workers. Worker i chooses effort

[10] It turns out that the restriction of β_i to lie between 0 and α_i is not needed for our purposes. We will return to this point in Section 4.

$e_i \in [0, e_{\max}]$, i = A, B. Department i's production p_i depends on the effort exerted by worker i in the following way:

$$p_i(e_i) = 4e_i - \frac{e_i^2}{100}, \quad i = \text{A, B.} \tag{2}$$

Effort is exerted at constant marginal costs $c_i \geq 0$ for Worker i. Worker i's material payoff, π_i, is then given by:

$$\pi_i(e_A, e_B) = s_i + \frac{1}{2}\Sigma_{j=A,B}\, p_j(e_j) - e_i c_i, \quad i = A, B. \tag{3}$$

From here onward, we consider the parameters $e_{\max}$=100, s_A=200, s_B=0, c_A=2, c_B=1 that we used in our experiment. Hence, A starts with a higher basic salary but faces higher marginal costs than B. These parameters ensure that, for any possible pair of effort levels up to the maximum of 100, $\pi_A \in [150, 350]$ and $\pi_B \in [0, 150]$, implying that Worker A always earns more than B (see online Appendix A for details). Therefore, when applying model (1) to the production game, for A only the guilt parameter β is relevant and for B only the envy parameter α.

More specifically (cf. online Appendix A), the IA model (1) yields the following dominant strategy for Worker A in the production game:

$$e_A = \begin{cases} 0 & \text{if } \beta_A \leq 0 \\ 200\beta_A & \text{if } 0 < \beta_A < 1/2 \\ 100 & \text{if } \beta_A \geq 1/2. \end{cases} \tag{4}$$

Similarly, B has a dominant strategy in the production game:

$$e_B = \begin{cases} 0 & \text{if } \alpha_B \geq 1 \\ 100(1-\alpha_B) & \text{if } 0 < \alpha_B < 1 \\ 100 & \text{if } \alpha_B \leq 0. \end{cases} \tag{5}$$

Therefore, if the IA model describes workers' preferences, A's effort choice in the production game is a dominant strategy that is a continuous function of the guilt parameter β, while B has a dominant effort level that is a continuous function of on her envy parameter α. As a consequence, equations (4) and (5) describe an equilibrium in dominant strategies for both a simultaneous version of the production game (where A and B simultaneously choose effort) and a sequential version of the game (where B sees A's effort choice before deciding herself). As discussed in the introduction, the above properties of the production game makes it particularly suited to test the predictive power of the IA model at the individual level and examine its robustness to reciprocity opportunities and stakes.

3. Experimental Design

The experiments were conducted at the […] laboratory of the […]. Subjects were recruited from the […] subject pool, which consists of approximately 2000 students, mainly […] undergraduates from various disciplines. 284 students participated and earned on average 38.60 euro, including a 7-euro show-up fee. All sessions lasted less than 60 minutes. At the start, participants are told that the experiment consists of several parts, and that the instructions to each part will be distributed before that part starts. Control questions are used to test understanding of these instructions. Subjects do not learn about their roles and their payoffs in any part until the end of the experiment. Because no information about others' choices is given until the end of the experiment, we consider each individual as an independent observation for our statistical analyses. Transcripts of instructions are presented in online Appendix C.

The experimental design exploits eight main treatments and two control treatments. All main treatments consist of four parts. Parts 1 and 2 measure each subject's social preferences (i.e., to obtain estimates of α and β) using three menus containing 10 dictator decisions each. In Part 3, subjects again make decisions in the two menus used in Part 2, but now for higher stakes. In Part 4, one of several variations of the production game is played. Using the parameter estimates from Parts 2 and 3, Part 4 allows us to test the predictive power of the IA model and its robustness to stakes and reciprocity opportunities.

The remainder of this section is organized as follows. We will continue by giving a more detailed overview of the four parts of the experiment and the treatments. Then, we discuss the details of the menus and the variations of the production game.

Overview of parts and treatments

The various parts of the experiment were computerized and programmed in z-tree (Fischbacher 2007). In Part 1, every subject is asked to make choices from a menu that we will use to measure efficiency concerns. As efficiency concerns are not the main focus of this paper, we discuss the details of this part, as well as the analysis of the results, in online Appendix D.

In Part 2, subjects are randomly assigned into pairs. Each subject is asked to make 10 choices in each of two menus, Menus 1 and 2. Each choice determines the subject's own payoff and the payoff of the subject she is paired with. At the end of the experiment one of the in total 20 decisions is randomly selected to be paid. In each pair one of the participants is appointed proposer and the choice of this proposer for the selected question is implemented.

In Part 3, Part 2 is repeated with higher payoffs. In three different treatments (varied between subjects), distinct payoffs are implemented. These are summarized in Table 1 (fourth column). Specifically, the payoffs of Parts 1 and 2 are multiplied by

10 (denoted by “low” in Table 1), 30 (“high”) and 60 (“highest”), compared to Part 2 (where the payoffs are labeled “lowest”). To control for income effects due to the possible earnings from the first two parts, each subject must choose whether or not to enter Part 3 (for a similar procedure, see Holt and Laury 2002). If a subject chooses to enter, she forfeits all earnings from the first two parts. If she chooses not to enter, all her previous earnings are kept, and she waits until this part finishes. In the experiment 234 subjects had to make this decision. Only 8 (3.4%) chose not to enter Part 3.

Finally, in Part 4, we randomly paired subjects and let them play one of two versions of the production game. In the first, subjects are not informed about their roles, and make their decisions simultaneously for both Worker A and B. We denote this simultaneous production game by SimProd. For each of the payoff levels of Part 3, we ran a “lowest” payoff simultaneous production game with payoffs in the order of magnitude of Parts 1 and 2 and a second payoff version with payoffs in the magnitude of Part 3 in the session concerned (i.e., either “low”, “high” or “highest”). These were varied between subjects. Varying the stakes in the production game allows us to test for stakes effects in the predictive power of the IA model.

In the second version of the production game, SeqProd, subjects make decisions in the production game sequentially (payoffs were scaled in the category “low”). For this, we again use a strategy method. Now, Worker B can condition her effort level on effort levels chosen by Worker A. For SeqProd, we also implemented two subtreatments, SingleRole and DoubleRole. Subjects know their roles (either A or B) in the SingleRole treatment, and only make decision for their own roles. In the DoubleRole treatment, subjects are not informed about their roles and need to specify an effort level as Worker A, and give a set of responding rules as Worker B as well. Only the decisions made for their true roles (revealed at the end of the experiment) is implemented. Comparing outcomes in SeqProd and Simprod allows us to isolate the effect of reciprocity opportunities on inequity aversion because in SimProd, Worker B cannot respond to Worker A's decision, while in SeqProd, she can.

In summary, there are six different versions of the production game, which were varied between subjects (see Table 2). The main distinction is between on the one hand the simultaneous version (SimProd), and on the other hand the two sequential versions (SeqProd/SingleRole and SeqProd/DoubleRole). The other versions varied the payoff stakes in SimProd. The distribution of the versions of the production game across sessions is shown in the final column of Table 1.

Finally, the last two rows of Table 1 show two control treatments (IX and X). The first, denoted by MenuOnly, contains only Part 3 (with "low" payoffs). This was used to check whether the experience in the low-stakes parts 1 and 2 has influence on subjects' decisions in the high-stakes menu tests. The other, denoted by ProdOnly, consists of only the (simultaneous) production game of part 4 (with "high" payoffs). This treatment was used to check whether having participated in the menu tests or not influences subjects' behavior in the production game.[11]

Choice Menus

We now describe the two choice menus used in parts 2 and 3 to measure social preferences.

[11] Average earnings were 12.45 in MenuOnly, and 38.79 in ProdOnly as compared to 41.30 across all other treatments. We do not find any evidence indicating that experience in the lowest-stakes menu tests (Part 1 and 2) affects decisions in the high-stakes tests (Part 3). A Kolmogorov-Smirnov test shows that the distribution of choices in MenuOnly does not significantly differ from that in Part 3 of the standard treatments, with the p-value of 0.992 for Menu 1 and 0.953 for Menu 2. Also, we do not find any evidence that the effort choices in the production game are influenced by experience in the previous parts. A Kolmogorov-Smirnov test does not reject the null hypothesis that the distribution of effort levels in ProdOnly are the same as in Part 4 of the other treatments (p-values are 0.245 and 0.685 for Worker A's and B's effort levels, respectively). We will therefore not further discuss these control sessions.

Menu 1

Menu 1 is used to measure the envy parameter α of the IA model in Part 2 of the experiment. This menu consists of 10 decisions (cf. Table 3). In each, the decision maker (denoted by 'proposer') is asked to choose between two options (A and B). Each option allocates money to the proposer and to an anonymous other participant (denoted by 'receiver'). For each of the ten decisions, the proposer is linked to the same receiver (though at most one decision will be selected for payment, as explained above). Each participant decides as if she is a proposer, because roles are not (randomly) determined until the end of the experiment.

In each of the payoff pairs in this menu, the proposer's payoff is lower than the receiver's, i.e., the proposer is always at the disadvantageous position (which is why it is informative about the envy parameter). As a consequence, the third term on the r.h.s. of (1) is equal to zero for all options considered in the menu. In each of the ten decisions, Option B gives 100 points to the proposer and 260 points to the receiver; and Option A is characterized by a lower (disadvantageous) inequality. Moving down from decision 1 to decision 10, the proposer's earnings decrease and inequality increases (see Table 3). From decision 4 onward, the proposer's own earnings are also lower in Option A than in B.

For decision 1, option A yields the proposer both higher own payoff and lower disadvantageous inequality than option B. Note that for non-negative α, both remaining terms on the r.h.s. of (1) then imply higher utility for A than for B. In fact, any proposer with $\alpha > -0.19$ (which would be all proposers under the standard IA assumption that $\alpha \geq 0$) will choose A. At the other extreme, consider decision 10. Here, choosing A means giving up 65 in own earnings (100−35) to decrease disadvantageous inequality from 160 (260−100) to 115 (150−35). Only individuals with strong envy ($\alpha \geq 1.44$) prefer A.

In this way model (1) determines for each decision question, a threshold for the envy level (α), above which Option A should be chosen and below which Option B should be chosen. This threshold is given in the last column of Table 3. If preferences are described by (1), a subject will switch when moving down from decision 1 to 10 at most once from choosing Option A to choosing Option B. It is easy to see that this switching point then identifies an interval for a proposer's envy level. More details on how subjects' envy levels are estimated can be found in online Appendix B.[12]

Menu 2

Menu 2 also consists of 10 decision questions (cf. Table 4), each containing an Option A and an Option B, with distinct payoff pairs for a proposer and receiver. In contrast to Menu 1, for all options the payoff of the proposer is higher than for the receiver. This means that all cases yield advantageous inequality for the proposer, which sets the second term on the r.h.s. of (1) equal to zero and allows us to use this menu to measure her guilt parameter β (cf. online Appendix B). Once again, each participant makes a decision as if she is a proposer because random role assignment is postponed until the end of the experiment.

Again, the payoffs for Option B remain constant across all 10 decisions, with the proposer earning 170, which is 120 more than the receiver (50). For the first decision,

[12] As mentioned above, previous studies have used a different approach to measure envy. They did so by eliciting responder rejection thresholds in ultimatum games. One difference with our approach is that reciprocity opportunities may play a role in ultimatum rejections. An anonymous referee pointed out a second difference. This is that our subjects do not have the extremely egalitarian outcome at their disposal where both players earn nothing. Under the linearity assumption in the IA model, this should not matter, but one cannot exclude that it will have a behavioral effect for boundedly rational decision makers.

Option A gives the proposer more (185) and yields lower inequality (95) than B. Any non-negative β then implies higher utility for A than for B. Moving down along the table, the own earnings in Option A decrease, as does the inequality. This increases the level of guilt needed to prefer Option A to B. The last column in Table 4 gives these threshold values for β.

Production Game

When introducing the production game to subjects, equations (2) and (3) are (of course) not used. Instead, participants are given a calculator to see the consequences of various effort levels by Workers A and B. Diagram 1 shows the computer screen used for this purpose. The screen is split in two halves, one for A's decision and one for B's decision.

In SimProd, we use the strategy method (with respect to a move by nature); not being informed of which role they will have, subjects are asked to make decisions as both Worker A and Worker B. For both roles' decisions, a subject can use the calculator to try out as many decisions as they like.

For each role, the participant can try out any effort levels by moving a scroll bar.[13] The table below the scroll bar shows the consequences of a decision for each worker. It shows the effort chosen, the 'bonus' (share of that department's production), the effort costs for the worker concerned, and the aggregate earnings for each worker. Note that the latter does not include the bonus to be earned from production in the other department. This is because that bonus depends solely on the other worker's efforts. The consequences of this other effort can also be tried out on the other half of

[13] Recall from eq. (3) that payoffs are linear in the production from both effort levels; there is no interaction effect between the two effort levels in the sense that the marginal effect of one's own effort on one's material payoff is independent of the other's effort.

the screen. After having practiced, the participant can choose a decision for both roles (A and B) and finalize by clicking an 'OK' button, after which she is asked to confirm her decisions.

In SeqProd, subjects make decisions in the production game sequentially. Now, Worker B can condition her effort level on effort levels chosen by Worker A. This is implemented as follows. Worker B chooses a set of 'responding rules' that consist of lower and upper bounds for effort chosen by A and a corresponding effort that B chooses for those efforts by A. B can formulate as many such rules as she wishes before finalizing her decision. An example is given in Diagram 2.

In this example, Worker B chooses effort level 1 if A chooses 8 or less, 13 if A chooses between 9 and 50 and 73 if A chooses 51 or more. The instructions in online Appendix C show how participants were informed about using this 'Decision Box'. Recall that we implement two version of SeqProd (cf. Table 2). In SeqProd-DoubleRole, each subject makes decisions as both Worker A (on the effort level) and B (on the responding rule). For this, we again use the strategy method. In SeqProd-SingleRole, each subject is informed about his/her role in the production game, and only decides for the role s/he is assigned to.

4. Results

This section starts with an overview of our estimates of the envy and guilt parameters for the IA model (1) as derived from the Menus 1 and 2. This includes a comparison to values estimated in previous studies. Then, we check the robustness of these estimates to the scaling of payoffs. Finally, we test the ability of the Fehr-Schmidt model to predict behavior in the production game. This enables us to test the robustness of the predictive power to reciprocity opportunities and stakes.

Estimates of Envy and Guilt

Our first estimates of the envy parameter (α) are derived from Menu 1. Specifically, the threshold values shown in the last column of Table 3 provide intervals for the estimated value. For example, a subject who chooses option A for decisions 1-4 and B for decisions 5-10 is estimated to have $\alpha \in [0.05, 0.16)$. Note that this procedure requires a maximum of one switch when moving down from decision 1 to decision 10. In fact, 216 of the 234 subjects (over 92%) who participated in a standard treatment session (i.e., all subjects except those in the MenuOnly and ProdOnly control sessions) are '(IA-)consistent' in this way, with the remaining 18 subjects labeled as "IA-inconsistent".[14] We will exclude the 18 IA-inconsistent from the further data analysis.

We have decided not to exclude subjects whose parameter estimates are inconsistent with the restriction $\beta \in [0, \alpha] \cap [0,1)$ proposed by Fehr and Schmidt (1999). Amongst the group of 216 IA-consistent subjects, 83 are estimated to have a negative value of α either in the lowest-stakes or the higher-stakes environment, which would indicate a preference for increased (disadvantageous) inequity aversion. However, as show in online Appendix D, all of these 83 subjects can be rationalized to have non-negative envy level by using a model that also allows for efficiency concerns.[15] In addition, we have decided not to exclude subjects with β estimates that contradict the restriction $\beta \notin [0, \alpha]$. In fact, many subjects violate this condition. 104

[14] Alternatively, one could make assumptions on preferences that aid in interpreting multiple changes. Given the ad hoc nature of such assumptions and the low number of subjects involved, we have decided not to do so.

[15] Note that previous studies like Blanco et al. (2011) could not identify negative alpha estimates: the games used predict the same choices for players having negative alpha as for those for whom alpha equals zero. Charness and Rabin (2002) and Engelmann and Strobel (2004) discuss how a preference for efficiency may give a subject reason to sacrifice own payoff for an increase in the payoff of another, even if this other already has a higher payoff. In an inequity aversion model like (1), such a subject may be perceived as having a negative envy level.

participants (48% of the IA-consistent subjects) have an estimated β level greater than their α.[16] Finally, one subject has a negative estimated β level, which is estimated to lie in the range (–0.60, –0.14].[17] We do not exclude any of these subjects as the IA model can straightforwardly be applied even if the assumption $\beta \in [0, \alpha] \cap [0,1)$ is not satisfied.

Our estimates of the IA-consistent subjects' envy are summarized in Figure 1(a). For comparison, we include the distributions reported by Fehr & Schmidt (1999) (FS henceforth) and Blanco, Engelmann & Normann 2011 (BEN henceforth).[18] Our results differ substantially from those previously found. χ^2-tests reject the null-hypotheses that the distribution of our estimates for α equals the distribution reported by FS or BEN, both at the 0.01 level. In fact, our estimates of subjects' envy parameters are almost completely (98.1%) clustered at the lowest interval ($\alpha < 0.25$). This stark difference between our estimates of envy and those in FS and BEN will be extensively discussed below.

In a similar way, we use Menu 2 to provide a first estimate of the guilt parameter, β. Our estimates are presented and compared to the FS and BEN estimates in Figure 1(b).[19] The distribution of our guilt parameter is more comparable to those reported by FS and BEN than envy. Again, however, our estimated distribution is more skewed

---

[16] This is close to the result reported by Blanco et al. (2011), that 23 (37.8%) of their 61 subjects violate the assumption that $\alpha \geq \beta$.
[17] In addition, 127 subjects are estimated to have a β in the range (–0.14, 0.11]. Using the midpoint of the interval as a point prediction would yield negative value –0.02. Note from Table 4, however, that these are subjects who switched from option A to option B at decision 3, which is the first decision where the own earnings are higher in B.
[18] The distribution given by FS has a few points with mass density (in particular, α=0, 0.5, 1, and 4). For comparison, we summarize our estimates in intervals of which the midpoints coincide with the FS estimates (except for the highest interval [1.25, +∞) which contains the FS estimate of 4). The intervals reported by BEN are different; here we followed the following procedure. First, we took their estimated values. Next, we used these to predict a switching point in our menus 1-2; then we processed these virtual switching points in the same way as the actual switching points in our data.
[19] Again, we center our intervals around the FS estimates, in this case β=0, 0.25, and 0.6.

towards the left. The null-hypotheses of the distribution of our β estimates being the same as the distributions estimated by FS and BEN, are both rejected at the 0.01 level.

Before analyzing the robustness of our estimates to stakes and testing the predictive power of our estimates, we briefly discuss their robustness to efficiency concerns, as analyzed in online Appendix D. Efficiency concerns are a confounding factor for the inequity aversion parameters as measured in the menus. It is relatively easy to see from Menu 1 that a downward bias would occur in the estimate of the envy parameter if an individual also cares about efficiency. Such a subject would choose the B option more often than someone without efficiency concerns because the sum of the payoffs are always greater under option B than under option A. As a consequence, decisions from Menu 1 alone would underestimate her α parameter. Similarly, the estimate of the guilt parameter is biased if we use Menu 2 in isolation for an individual who also has efficiency concerns. Here the bias is non-monotonic in the actual value of β. Up to decision 6, the sum of the payoffs is higher in option A. Therefore, subjects who care about efficiency and have a guilt parameter below 0.47 may choose A more often than those with the same parameter who do not care about efficiency. This will cause an upward bias in the estimated β. The reverse is true for guilt parameters greater than 0.53.

In online appendix D, we correct for these biases by jointly estimating envy, guilt, and efficiency concerns using the choices in Menus 1 and 2, and a third menu introduced for this purpose. Our estimates of IA parameters turn out to be quite robust to allowing for efficiency concerns. First, we compare the derived distributions of α and β across the categories used in Figure 1, for the cases with and without correction for efficiency concerns. χ^2 tests give a p-value of 0.250 for α and 0.767 for β, so we cannot reject the null hypothesis that the distributions of the IA estimates using

Menus 1 and 2 alone are the same as the estimates derived when correcting for efficiency concerns by using the third menu as well. The FS categorization is very crude, however. In online appendix D, we also present a more fine-tuned comparison between intervals of estimates for α and β using model (1) and a more general model that includes efficiency concerns. We observe that 80-90% of the estimated α (or β) values that are consistent with the IA model based on Menu 1 (or 2) are also consistent with the general model jointly based on Menus 1 and 3 (or 2 and 3).

Engelmann (2012) elegantly analyzes how the IA model could be extended to incorporate efficiency concerns. He illustrates that a general linear model with distinct parameters for envy, guilt, and efficiency can be fully captured by a two-parameter IA model, unless one simultaneously considers games with different numbers of players. This motivated our choices underlying the analyses in Appendix D. Another implication of Engelmann's (2012) analysis is that if the number of players is fixed, predictions using the IA model are the same even if efficiency concerns are present. In other words, the IA model's predictive power for behavior in the production game does not improve if efficiency concerns are taken into account as in both the menus and the production game, the number of players is two. For this reason, we can use our estimates from Menus 1 and 2 to derive predictions for the production game, irrespective of whether efficiency concerns play a role.

Robustness to Stakes

Eight out of the 216 IA-consistent subjects chose not to enter Part 3. Therefore, we have 208 observations for which we can compare α's and β's estimated under different payoff scales. To start, Figure 2 shows the estimated distributions for the various payoff scales used. Recall from Table 1 that this is a within-subject comparison: each subject participated in the benchmark scale ("lowest") and in one of the higher scale treatments.

A first impression from the figure is that estimates of the envy parameter are insensitive to changes in the stakes, though this is likely related to the almost complete lack of envy in the first place. For guilt, there appears to be a shift towards lower β-values with increasing scale. To test whether this effect is statistically relevant, we use Wilcoxon sign-rank tests comparing at the individual level the IA parameters obtained from the two scale menus an individual participated in. The results are summarized in Table 5.

These tests only reject the null hypothesis of no scale effect when payoffs are multiplied by 60 ("highest"). The envy parameters obtained from "highest" are found to be significantly higher than from "lowest".[20] For the estimates of the envy parameter for lower scales, the test results do not reject the null hypothesis that they are invariant to changes in stakes. Therefore, we do not reject the hypothesis that the disutility from the payoff difference is linear in disadvantageous inequality for intermediate stakes. Similarly, the estimates of the guilt parameter are also robust to payoffs being scaled up by a factor 10 or 30. We again reject the null of no stakes effect when payoffs are 60 times higher than in the benchmark case (i.e., in "highest"). This result implies that subjects feel less guilt about receiving more money than others when the amount earned is (much) more. In other words, the marginal disutility of advantageous inequity is decreasing in income. Again, the linearity assumption for advantageous inequity is warranted for moderate increases in payoff

[20] Part of this change can be attributed to 12 subjects with negative α estimates in "lowest" showing less negative or zero envy levels in "highest". Moreover, we note that of the 50 observations, 36 (72%) have the same estimate for the envy parameter in the lowest and highest stakes. The significance for the Wilcoxon statistic stems from 12 of the remaining 14 subjects exhibiting higher envy when the stakes are highest.

levels, but not for major 60-fold increases. This finding suggests that the IA model's predictive power may diminish if the stakes vary greatly between the environment where IA preferences are measured and the environment of interest. We will return to this issue when discussing the IA model's predictive power in the production game.

Behavior in the Production Game

We will use the production game to test the predictive power of the IA model. Before doing so, we note that this is a novel game that has not previously been studied in the laboratory. We therefore start with a brief overview of observed behavior in this game. Table 6 gives an overview of average effort choices in the various treatments of this game.

The results for the simultaneous version of the game (SimProd) show that A-workers exhibit less effort than B-workers. Across all payoff levels, the average (weighted by the number of observations) efforts are 41.8 and 95.7, respectively. It appears that in the simultaneous game, subjects take into account the distinct marginal costs of effort (2 for A and 1 for B). Because the benefits from production are shared equally and A has lower aggregate costs, the average production choices increase the inequality between A and B (recall that A always earns more than B) compared to the case where they only have their basic salaries. The distinct payoff scales in the game yield only small differences, with the exception of the highest payoffs (60 times the lowest). We use two-sided Kolmogorov-Smirnov (KS) tests to investigate whether the distribution of efforts (separately for A and B) in "lowest" differs significantly from that in "low", "high" or "highest". This is not the case for "low" or "high" or for B's effort in "highest" (all $p>0.166$). Worker A produces significantly less effort in "highest" than in "low", however (p=0.001). Hence, only an increase in payoff scale with a factor of 60 has an effect on (A's) effort choices.

The differences between the two workers do not appear in the two sequential games. When B can condition her effort on that of A, A produces much more effort and B much less than in the simultaneous case. Comparing to the simultaneous case with the same "low" payoff scale, the increase for worker A is significant for the SingleRole case (KS, p=0.029) and insignificant for DoubleRole (KS, p=0.479). The decrease compared to SimProd in worker B's effort is significant for SeqProd/DoubleRole (KS, p=0.022) but insignificant for SeqProd/SingleRole (KS, p=0.109). All in all, the possibility of explicit reciprocation introduced by the sequential nature of the game thus destroys the inequity increasing effects observed in the simultaneous game by inducing B-workers to provide less effort and A-workers to provide more.

Next, we test whether using the strategy method has an effect on the effort levels chosen. We do so by comparing the SingleRole and DoubleRole treatments. It turns out that the distributions of effort levels are not significantly different (KS; p=0.243 for worker A and p=0.725 for worker B). Therefore, in the remainder of the analysis, we will not distinguish between the two treatments.

Finally, we consider whether the 'responding functions' that B-workers submitted in the SeqProd treatments exhibit reciprocity. This is the case if their effort choices are increasing in A's effort. We tested this in the following way. First, for each IA-consistent worker B in the SeqProd treatments (13 SingleRole and 26 in DoubleRole), we create 101 fictitious A decisions (choosing effort levels 0,1,…,100). Then we used Bs' responding rules to determine for each B her effort in response to each of these 101 effort levels. Finally, we regressed B's effort on A's (fictitious) effort. This shows that Worker B on average increases her effort by 0.374 for each unitary increase in A's effort (with an associated p-value <0.001). We conclude that B-workers on average do reciprocate A's effort.

Predictive Power of the IA model in the Production Game

Predictive power of Part 2 estimates

We can use the estimates of an individual's α and β to predict her behavior in the production game. We do so in two ways. First, we use the fact that the models predict specific relationships between individual envy and guilt parameters on the one hand and their effort choices on the other. We use regression equations to investigate whether we can reproduce these relationships. Second, we will use individual subjects' estimated parameters to derive for each subject an interval in which her effort level is predicted to lie and check whether her observed effort level indeed lies in this interval.

We start by presenting the results from regressions of effort choices on individual envy and guilt parameters. Recall that we have two versions of the production game, SimProd and SeqProd (cf. Table 2). To test whether the predictions for the production game are supported by the observations from the lab, we ran the following regression:

$$e_{R,i}^{T} = \delta_{R,0}^{T} + \delta_{R,1}^{T}\alpha_i^{IA} + \delta_{R,2}^{T}\beta_i^{IA} + \varepsilon_{R,i}^{T} \quad (6)$$

where T = SimProd, SeqProd, $R = A, B$ represents the subject's role in the production game,[21] and i is the index of a subject. α_i^{IA} and β_i^{IA} are the IA estimates of i's envy and guilt parameter (as derived from the lowest-stakes Menus 1 and 2 in Part 2),

[21] Recall that in most treatments, subjects played both roles.

respectively. Table 7 gives the estimated coefficients of the δ's in (6) and shows the predictions from the IA model (eqs. (4) and (5)).

Comparing the results in Table 7 to the theoretical predictions of the IA model shows that in the simultaneous production game (SimProd) the relationships between envy/guilt and effort go in the direction of the predicted coefficients. Worker A's effort level only relies significantly (and positively) on her guilt parameter. The estimated coefficient is below the predicted level of 200, however. Worker B's effort is strongly affected by the envy parameter, which has the predicted negative impact on her effort level, though again the marginal effect falls short of what is predicted. Moreover, in contrast to what theory predicts, Worker B's effort is also affected by the guilt parameter, albeit modestly and marginally significantly. For the two treatments of the sequential production game, the regression outcome differs substantially from the theoretical prediction, however. We observe no significant evidence in support of any of the predicted effects.

As a second way to investigate the predictive power of the IA model, we directly compare for each subject the predicted and observed effort levels. Using the estimated interval of the envy and guilt parameters derived from each subject's choices in Menus 1 and 2, eqs. (4) and (5) yield for each subject a prediction for the interval in which her effort as Worker A or B will lie. We derive these intervals for each subject and simply check whether her observed efforts in each worker role lie within the predicted intervals.[22] The results are presented in Table 9, below. The aggregate results for SimProd show that 53.9% (90 out of 167) of the observed efforts as Worker A and 87.4% (146 out of 167) as Worker B lie in the interval predicted by the

[22] This method does not correct for the length of the intervals, but note that we have 11 intervals for each parameter, which makes the average length much shorter than in the often-used FS categorization.

IA parameters estimated from the lowest-level menus 1 and 2. The success rate is much lower in SeqProd, where only 22.2% (8 out of 36) of Worker A's efforts and 50.0% (14 out of 28) of Worker B's efforts lie in the predicted intervals.[23] We will return to this difference between the two games when discussing our results below.

Robustness of predictive power to variations in stakes

To test whether the stakes influence the predictive power of the IA model, we run regression (6) again using α_i^{IA} and β_i^{IA} estimates derived from the menus where the stake levels coincides with those used in the production game the player concerned participated in (Menu 1 and 2 of Part 2 in treatments I, III, and V and the high-stake menus of part 3 in treatments II, IV, VI, VII, VIII; see Table 1). It turns out that with this adjustment, the regression of (6) generates qualitatively the same results (Table 8) as using the lowest-stake IA estimates. Envy and guilt are still estimated to have significant influence on worker B/A's effort levels, respectively, in SimProd. Also, in line with Table 7, the new regression shows no significant effects in SeqProd, except that the negative effect of guilt on Worker B's effort is estimated to be significant at the 5% level, which remains inconsistent with the theoretical prediction. Further support for the conclusion that the stakes do not matter for these estimates stem from the observation that the levels of the estimated coefficients are not significantly closer to the theoretical predicted levels than those estimated when using the lowest-stake IA estimates.

We also test the success rate of individual predictions for the production game by using the IA parameters estimated from the same-stake menus (i.e. the menus in Part

[23] For SeqProd, the data on Worker A's effort are from all IA-consistent subjects in SeqProd/DoubleRole and those who have been assigned the role of Worker A in SeqProd/SingleRole. The data used for Worker B are selected in an analogous way.

2 if the production is using the lowest stake, and those in part 3 otherwise). The overall results are almost the same as those obtained when using the lowest-stake IA estimates. At an aggregate level for SimProd, out of the 166 observations, 92 (55.4%) of Worker A's and 144 (86.8%) of Worker B's effort levels lie in the predicted intervals (cf. the penultimate row of Table 9). For SeqProd, the same-stake prediction of Worker A/B's efforts is correct 22.6% (7 out of 31) and 52.0% (13 out of 25 obs.) of the time, respectively. This shows that adjusting the stakes used to estimate envy and guilt to the level used in the game under analysis does not significantly improve the predictive power of the IA model.

Note that the case closest to extrapolation of results from laboratory experiments might be the case in treatment VI where estimates from Part 2 (a typical laboratory level of stakes) are used to predict behavior in a highest-stake production game (reflecting that stakes in the field are often much higher than in the laboratory). Here, we observe a successful prediction in 68.2% of the cases for worker A and in 86.4% for worker B and that for both worker types, the predictive power does not improve if the estimates from part 3 (where the participants' inequity aversion parameters are measured under the highest stakes) are used instead of the estimates from Part 2.

Robustness of predictive power to reciprocity opportunities

The weak predictive power of the IA model in SeqProd (observed irrespective of whether the parameters are estimated from Part 2 or Part 3) strongly suggests that the possibility of explicit reciprocation has an impact on decision makers' preferences for envy and guilt. Therefore, we conclude this section by presenting direct estimates of individuals' envy (α) and guilt (β) parameters derived from decision in SimProd and

SeqProd. We use the inverse relationships of (4) and (5) to do so.[24] The resulting distributions are displayed in Figure 3. For comparison, we include our estimates from Menus 1 and 2 as well as the FS and BEN distributions. For SeqProd (both the DoubleRole and SingleRole treatments) worker B gives responding rules instead of a specific effort level. To obtain a single choice, we again use the effort level that is realized after the chosen responding rule has been applied to A's chosen effort level. As a consequence, only observations from subjects who are assigned to be worker B in SeqProd (28 observations in total) are used to estimate α; in contrast, all effort levels by workers A (also those that ex post were allocated to be a worker B) are used to estimate β.

Figure 3(a) shows that the simultaneous production game yields a distribution of envy that is very much like the distribution estimated with the IA model from the choices for Menu 1 (cf. Figure 1(a)). Hence, in SimProd, where explicit reciprocity is ruled out, the envy parameters are very similar to those estimated from Menu 1 (where reciprocating the other's choice is not feasible either). This explains why the regression results reported for SimProd in Table 7 are quite consistent with the theoretical prediction.

The results are different for the sequential production game, where explicit reciprocity is possible. Here, far fewer subjects (67.9% as opposed to 94.0% in SimProd) are estimated to have an envy parameter in the lowest category. We observe a substantial number of subjects showing envy levels at higher values. This suggests that high envy levels are triggered by (negative) reciprocity. As a result, the distribution we derive from SeqProd is closer to those reported in FS and BEN. Recall

[24] This way of obtaining estimates of the model's parameters is the method traditionally used (starting with Fehr and Schmidt 1999 themselves). As explained above, the production game is better suited for this purpose than the games used in previous studies.

that their envy parameters were also derived from an environment where subjects may be exposed to negative reciprocity.

In Figure 3(b), the distribution of guilt levels derived from Worker A's effort, in SimProd and SeqProd, are compared to those in FS and BEN. Similar to envy, we observe higher guilt levels in SeqProd than in SimProd. This may be attributed to a strategic choice by A, anticipating B's negative reciprocal response if her effort level is too low.

5. Concluding Discussion

In this paper, we have examined the robustness of parameter estimates and the predictive power of the IA model. Our main findings are the following. First, our estimates of disadvantageous inequity aversion (envy) and advantageous inequity aversion (guilt) are reasonably robust to increases in the stakes (though inequality seems to become relatively less important when payoffs are scaled up very strongly). Second, while our results indicate that guilt is important for some subjects we find far less evidence of envy than has been observed in previous studies. Only when explicit reciprocity is possible, do we observe more subjects for whom envy plays a role. Third, we observe that in settings where explicit reciprocity is ruled out, the model's predictive power is higher than what results from previous studies suggest.

The effect of explicit reciprocation in our production game shows up in changes in both workers' behavior in the sequential production game compared to its simultaneous counterpart. In the sequential game, the second-mover may condition her choices on her perception of the opponent's kindness or meanness, which may trigger her motivation to reciprocate or retaliate observed choices, something that is impossible in the simultaneous game. In other words, the second-mover may have reciprocal preferences that are distinct from her feelings of envy or guilt (see also Falk and Fischbacher 2006 and Charness and Rabin 2002, pp. 824-825). If this is the case, these will surface in the sequential game and choices made there may be mistakenly interpreted as evidence of inequity aversion. Similarly, the first-mover may behave strategically differently in the sequential game than in the simultaneous game, anticipating changes in the second-mover's decisions in the former case.

Fehr and Schmidt (1999) acknowledge the possibility that the parameters of their model can be interpreted in two ways, when they argue that "positive α_i's and β_i's can be interpreted as a direct concern for equality as well as a reduced-form concern for intentions. […] As a consequence, our preference parameters are compatible with the interpretation of intentions-driven reciprocity." In our view, strategic and reciprocal tendencies should be distinguished from inequity aversion per se, however. In other words, we favor the view that reciprocal preferences should be distinguished from preferences with respect to equality. The alternative (that preferences about inequality vary with the environment, i.e., whether or not explicit reciprocity is possible) requires allowing for endogenous preferences, which would substantially reduce the predictive power of the model. In our preferred interpretation, a comparison between the estimates of envy and guilt estimates from the simultaneous and the sequential production games reveals that in an environment with explicit reciprocity, subjects' behavior will yield higher estimates of inequity aversion. Hence, the current literature (where estimates of the envy parameter are traditionally derived from responder behavior in the ultimatum game) may provide biased estimates of the envy parameter.

In short, we believe that the control for explicit reciprocity offered by our design (both in the choice menus and in the production game) is important for isolating preferences for equality and therefore for accurately measuring pure inequity aversion levels. We believe that our design provides stronger evidence for the robustness and predictive power of the IA model than was found in any previous research. Our understanding of this finding is that the IA model does a good job in explaining and predicting subjects' behavior in environments where explicit reciprocity is ruled out. However, when explicit reciprocity may play a role, the model should be augmented with a reciprocity term (e.g., as in Charness & Rabin 2002). Much of the previous literature seems to have taken an alternative 'as if' approach, in the sense that any choice that simultaneously yields lower inequity and lower own payoff is reflected in the measured inequity aversion parameters, irrespective of whether other motivations could be involved. Whether or not this is a problem in practice depends on one's goals. If one is interested in applying the model to an environment where explicit reciprocity is deemed to be important, one can measure inequity aversion in a situation that also allows for reciprocity and interpret the resulting parameters 'as if' they measure envy and guilt *per se*. Though one may question the interpretation of the parameters of the model, the model's predictive power need not be affected. Nevertheless, from a scientific point of view the distinction between various kinds of preferences seems important.

ACKNOWLEDGMENTS

The authors are very grateful for valuable comments by Dirk Engelmann and Ernst Fehr on an earlier draft. Participants at the ESA 2011 conference in Chicago IL, CeDEx-CBESS-CREED 2011 workshop in Nottingham, UK, and NYU-CESS/CREED 2011 workshop in New York NY also provided useful suggestions.

References

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., and List, J. A. 2011. "Stakes matter in ultimatum games." *American Economic Review*, 101(7): 3427-3439.

Binmore, K. and Shaked, A. 2010. "Experimental economics: Where next?" *Journal of Economic Behavior & Organization*, 73 (1): 87-100.

Blanco, M., Engelmann, D., Koch, A.K., and Normann, H.T. 2014. "Preferences and beliefs in sequential social dilemma: a within-subjects analysis," *Games and Economic Behavior*, 87: 122-135.

Blanco, M., Engelmann, D., and Normann, H.T. 2011. "A within-subject analysis of other-regarding preferences". *Games and Economic Behavior*, 72: 321-338.

Bolton, G.E. and Ockenfels, A. 2000. "ERC: A theory of equity, reciprocity, and competition". *The American Economic Review*, 90 (1): 166-193.

Brandts, J., Fatás, E., Haruvy, E. and Lagos, F. 2014. "The impact of relative position and returns on sacrifice and reciprocity: An experimental study using individual decisions". Forthcoming, *Social Choice and Welfare*.

Brandts, J., and Solà, C. 2001. "Reference points and negative reciprocity in simple sequential games." *Games and Economic Behavior*, 36(2): 138-157.

Camerer, C. 2003. "Behavioral game theory: Experiments in strategic interaction". Princeton University Press, Princeton.

Carpenter, J., E. Verhoogen, and S. Burks. 2005. “The Effect of Stakes in Distribution Experiments.” *Economics Letters* 86:393-398.

Charness, G. and Dufwenberg, M. 2006. “Promises and Partnership”. *Econometrica*, 74(6): 1579-1602.

Charness, G. and Rabin, M. 2002. “Understanding social preferences with simple tests”. *The Quarterly Journal of Economics*, 117(3): 817-869.

Cooper, D. and Kagel, J.H. 2012. “Other regarding preferences: a selective survey of experimental results”. Forthcoming in Kagel, J.H., and Roth, A., *The Handbook of Experimental Economics, Vol. 2*. Princeton University Press, Princeton.

Dannenberg, A., Riechmann, T., Sturm, B. and Vogt, C. 2007. “Inequity aversion and individual behavior in public good games: an experimental investigation”. Working paper.

Engelmann, D. 2012. “How not to extend models of inequality aversion”. *Journal of Economic Behavior & Organization*, 81(2): 599–605.

Engelmann, D. and Strobel, M. 2004. “Inequality aversion, efficiency, and maximin preferences in simple distribution experiments”. *The American Economic Review*, 94(4): 857-869.

Falk, A., Fehr, E., and Fischbacher, U. 2003. “On the nature of fair behavior.” *Economic Inquiry*, 41(1): 20-26.

Falk, A., Fehr, E., and Fischbacher, U. 2008. “Testing theories of fairness – Intentions matter.” *Games and Economic Behavior*, 62: 287-303.

Falk, A. and Fischbacher, U. 2006. “A theory of reciprocity”. *Games and Economic Behavior*, 54(2): 293-315.

Fehr, E., Naef, M. and Schmidt, K.M. 2006. “Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment”. *The American Economic Review*, 96(5): 1912-1917.

Fehr, E. and Schmidt, K.M. 1999. “A theory of fairness, competition, and cooperation”. *The Quarterly Journal of Economics*, 114(3): 817-868.

Fehr, E. and Schmidt, K.M. 2010. “On inequity aversion: a reply to Binmore and Shaked”. *Journal of Economic Behavior & Organization*, 73(1): 101-108.

Fehr, E., Tougareva, E., and Fischbacher, U. 2014. “Do high stakes and competition undermine fair behaviour? Evidence from Russia.” *Journal of Economic Behavior & Organization*, 108: 354-363.

Fischbacher, U. 2007. “z-Tree: Zurich toolbox for ready-made economic experiments.” Experimental Economics, 10(2): 171-178.

Goeree, J.K. and Holt, C.A. 2000. “Asymmetric inequality aversion and noisy behavior in alternating-offer bargaining games”. *European Economic Review*, 44(4-6): 1079-1089.

Holt, C.A. and Laury, S.K. 2002. “Risk aversion and incentive effects”. *The American Economic Review*, 92(5): 1644-1655.

List, J.A. and T. Cherry. 2008. “Examining the Role of Fairness in High Stakes Allocation Decisions,” *Journal of Economic Behavior and Organization*, 65(1): 1-8.

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

Teyssier, S. 2012. “Inequity and risk aversion in public good games,” *Public Choice* 151(1): 91-119.