

# Inequity Aversion Revisited

Yang Yang

Sander Onderstal

Arthur Schram

*University of Amsterdam and Tinbergen Institute, The Netherlands*

October 4, 2012

## **ABSTRACT:**

We provide the first systematic study of the robustness of parameter estimates for the Fehr-Schmidt (1999) model of inequity aversion with respect to (i) the occurrence of efficiency concerns; (ii) the scale of payoffs; and (iii) the game used (i.e., cross-game consistency). Moreover, we provide evidence of a bias in the estimates that occurs if one does not correct for strategic considerations and reciprocity. Our results show that the model is remarkably robust, but that previous estimates (especially of the disutility of disadvantageous inequity aversion) may overestimate the importance of inequity aversion plays.

**KEYWORDS:** Inequity aversion, efficiency, reciprocity, robustness

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

Please address correspondence to: Yang Yang, CREED, Amsterdam School of Economics, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands.

Email:

Yang: *Y.Yang@uva.nl*

Onderstal: *Onderstal@uva.nl*

Schram: *Schram@uva.nl*.

## **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. Financial support from the University of Amsterdam Research Priority Area in Behavioral Economics is gratefully acknowledged.

## 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) and Bolton and Ockenfels (2000) would have said it all.<sup>1</sup> In our view, this is not the case. We believe that this extensive literature has at least two important gaps. First, a systematic test of the model's robustness to systematic variations in its premises has not been undertaken. Second, the literature has failed to adequately distinguish between inequity aversion *per se* and (preferences for) reciprocity. This paper aims to fill these gaps. This is important, because both aspects pertain directly to the empirical applicability of the model. Given that the IA 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 approach to its strengths and weaknesses seems of utmost importance.

Such a '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 main focus is on the methodology used by Fehr and Schmidt, however. Though we agree that a critical assessment of methods is important, it is also necessary to stress test a model's robustness, 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 May 2012 this had increased to over 6000). Only a few of these (to be reviewed below) investigate the model's predictive power, none systematically studies its robustness. This is the issue that this paper addresses.<sup>2</sup>

In particular, we will study three aspects of the robustness of the IA model's estimated parameters, to wit, (i) to the inclusion of efficiency concerns; (ii) to variations in payoff scales; (iii) to the game concerned (i.e., cross-game consistency). Moreover, we will test whether the possibility to reciprocate others' choices affects the extent of inequity aversion.<sup>3</sup> To do all of this, we will apply laboratory

---

<sup>1</sup> See Cooper and Kagel (2009) for a review of the literature on inequity aversion.

<sup>2</sup> 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.

<sup>3</sup> Formally, our test of the role of reciprocity may also be interpreted as a robustness check of the IA parameters. In this paper, we prefer to view IA with and without reciprocity as separate models, however, because reciprocity directly affects the interpretation of the IA parameters. This is discussed in our concluding section.

experiments, as many studies on IA have done before us. We will introduce simple choice menus that allow us to directly measure each individual's level of IA. Moreover, we will introduce a new game, which we call the 'production game'. This has various desirable properties (to be discussed below) that will allow us to straightforwardly test the IA model's predictions.

We use a two-step experiment to test the IA model. The first step is to estimate each subject's IA preferences by using a set of choice menus. We estimate such preferences using two different models, one with and one without efficiency concerns, and check the robustness to the IA model by comparing the two estimates. Moreover, we check their robustness to scaling (i.e., we test the linearity assumption underlying the model) varying the payoff scales in the menus. In the second step, we let subjects play the production game. By comparing their decisions in this part to the theoretical prediction derived from the estimates of their individual IA levels as obtained in the first step, we test their robustness to the game concerned.

Of course, there have been other attempts to test the IA model. In fact, economists have conducted various laboratory experiments to test the IA model and, typically, to compare it to other social preferences models (for example, see Charness and Rabin 2002, Engelmann and Strobel 2004, and Brandts et al. 2011). Typically, such studies develop a series of simple games for which various models (including IA) 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 IA (though the latter result has been disputed by Fehr et al. 2006). Notwithstanding the elegance and usefulness of such horse races between models, they do have drawbacks. For one thing, they are ultimately only informative about the models concerned and for the games chosen. Moreover, not the models themselves are tested, but their (comparative) predictions. This is different in our approach. We will directly test the model's premises with respect to preferences.

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 two drawbacks. First, because

the optimal strategy for the players in several of the games used depends on the extent to which the opponents care about inequality, the validity of the test is based on an assumption that subjects hold correct beliefs on the distribution of IA preferences in the population. This seems quite demanding, especially in one-shot games. Second, 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) test predictions at the individual level i.e., without the payoffs being dependent on other individuals' decisions. Such an individual-level test can directly check the within-subject consistency for each individual without requiring any assumptions on the rationality of beliefs. Our method is also an individual-level test of the IA model.

Aside from directly measuring IA in various circumstances (as opposed to comparing its predictions in a horse race with other models), there are other advantages to our approach, compared to previous individual-level tests of the IA model. First, we consider efficiency concerns as an addition to the IA model, so that the original model is nested in the extended version with such concerns.<sup>4</sup> This allows us to directly test the robustness of measured IA parameters to allowing for such efficiency concerns. Second, we introduce the production game, in which a player's behavior predicted by the IA model is a continuous function of her IA levels. In the games traditionally used to test the model (e.g., the prisoners' dilemma game 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. Third, subjects' risk attitudes are irrelevant in our experimental environment. This is contrast to most of the games traditionally used for the analysis of social preferences (e.g., the public good game, prisoners' dilemma, or ultimatum game; see Fehr and Schmidt 1999, Blanco et al. 2011). These traditional games are characterized by strategic uncertainty; hence, a subject's decisions may be affected by her risk attitudes, which could yield biased estimates of her IA level.<sup>5</sup> In contrast, none of the decisions used in

---

<sup>4</sup> We are aware of Engelmann's (2012) critical discussion of introducing efficiency concerns in this way. We will discuss the relevance of this criticism below, when introducing the theoretical model.

<sup>5</sup> 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.

our experiments involve any strategic uncertainty. Therefore, risk attitudes play no role, allowing for a more accurate test of the IA model. Finally, we will present two versions of the production game that only differ in the simultaneity of decisions. A comparison will allow us to isolate the role of ‘explicit reciprocity’ (Charness and Rabin 2002) from IA-preferences in a way not previously done.

The approach that is probably closest to ours is the one applied by Blanco et al. (2011).<sup>6</sup> 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. Also, their method to measure the individual guilt level is based on a modified dictator game, which has a structure similar to the test menu we use to measure advantageous inequity aversion (the disutility derived from feeling guilty about having a higher payoff than others).

There are three main differences that distinguish our work from Blanco et al. (2011), however. The first is that we test the IA model’s robustness to allowing for efficiency and variations in payoff scales, which we believe are two important issues in the specification of the model. Second, the methods we use to measure individual IA levels and to test the internal consistency of the IA model are different from Blanco et al.’s (2011). Instead of deriving the disadvantageous IA parameters indirectly from the ultimatum game, we use a set of simple menus to directly measure IA preferences (not only the guilt parameter). Third, instead of testing the IA model using classic games such as public good games and two-stage prisoner’s dilemmas we introduce a novel game, the production game, for this test. As we will see, a major advantage of our methods is that several important factors (e.g., reciprocity concerns and risk attitudes) that may affect behavior and may play important roles in the games used in Blanco et al. (2011) are well-controlled here. As will be shown in Section 4, our results differ from those reported in Blanco et al. (2011) indicating the importance of controlling for such factors.

Our finding shows, first, that the estimates of the inequity aversion model are robust to the inclusion of efficiency in the model and –to a large extent– also robust to variations in the payoff scales. Second, the IA model predicts individual subjects’ behavior in the production game quite well. Adding the possibility of reciprocity to the game strongly reduces the model’s predictive power, however. Finally, our

---

<sup>6</sup> Dannenberg et al. (2007) present an application of the Blanco et al. method.

estimate of the distribution of the guilt parameter is roughly similar to the one reported in previous studies, including Fehr and Schmidt (1999), but we find much lower levels of the envy parameter (measuring disutility of disadvantageous inequality). 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. All in all, we conclude that distributional preferences measured by the IA model are remarkably robust but that previously measured disadvantageous inequity aversion is strongly influenced by negative reciprocity. 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.

## 2. Theory

We will derive hypotheses using the classic Fehr-Schmidt (1999) model of inequity aversion and an extension of this model that includes efficiency concerns. In the current section, we present these models. Moreover, we introduce here the production game that we will use below and analyze it in the light of the two models.

Consider a population of  $n+1$  individuals. In the Fehr-Schmidt (1999) 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:

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

where  $(x, y_1, y_2, \dots, y_n)$  denotes a payoff bundle,  $x$  is a player's own payoff, and  $y_j, j=1, \dots, n$ , denote the other  $n$  players' payoffs.  $\alpha_i \geq 0$  is an envy parameter (Engelmann and Strobel 2004) measuring the marginal disutility of disadvantageous inequality.  $\beta_i$  measures the marginal disutility (guilt)<sup>7</sup> related to advantageous inequality, with  $\beta_i \in [0, \alpha_i] \cup [0, 1)$ .<sup>8</sup>

One of our robustness tests will investigate the effects of allowing for preferences for efficiency. In a simple model, (positive) utility is derived from own earnings as well as from aggregate earnings (i.e., efficiency):

Efficiency model:

$$U_i(x, y_1, y_2, \dots, y_n) = x + \gamma_i(x + \sum_j y_j), \quad (2)$$

where  $\gamma_i \geq 0$  measures the marginal utility of aggregate earnings (efficiency).

For the robustness test, we will use a general model that incorporates both inequity aversion and efficiency concerns.

General model:

$$U_i(x, y_1, y_2, \dots, y_n) = x - \alpha_i \frac{\sum_j \max\{y_j - x, 0\}}{n} - \beta_i \frac{\sum_j \max\{x - y_j, 0\}}{n} + \gamma_i(x + \sum_j y_j), \quad (3)$$

where  $\alpha_i, \beta_i, \gamma_i$  still satisfy the above mentioned assumptions in the IA and efficiency models.<sup>9</sup>

The incorporation of efficiency concerns in a general model like (3) is elegantly analyzed in Engelmann (2012). The author illustrates that such a general model with

---

<sup>7</sup> 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).

<sup>8</sup> It turns out that the restriction of  $\beta_i$  to lie between 0 and  $\alpha_i$  is not needed for our purposes. We will return to this point in Section 4.

<sup>9</sup> Note that our formulation of efficiency concerns in (3) implies that player  $i$ 's utility increases with  $n$ . This means, for example, that  $i$  is better off if more people join her reference group (with the same payoffs as hers). This reflects our interpretation of efficiency concerns. Alternatively, one could consider a formulation such as  $\gamma_i \frac{(x + \sum_j y_j)}{n+1}$ , where only the average earnings per reference group

member affect  $i$ 's utility. In our view, (3) provides a more intuitive description of efficiency concerns. Considering the following two situations, for example. In one, an individual may give up 1 dollar to increase average income by 0.5 in a group of four; in the other, she faces the option to pay 1 dollar to increase the average income by 0.5 for a group of 1000 people. Though this is, in the end, an empirical question, we believe that one would likely make different choices in these two situations. The alternative formulation of efficiency concerns would predict the same choice in both cases because it fails to capture the fact that people care not only about social average income, but also about the number of people in their reference group. This consideration underlies our choice of the formulation of efficiency concerns in (3).

distinct parameters for envy, guilt and efficiency, can be fully captured by a two-parameter inequity aversion model, unless one simultaneously considers games with different numbers of players. It will become clear below that our experiment does vary the number of players.<sup>10</sup> An implication of disregarding efficiency concerns is then, that when directly measuring the two IA parameters in the IA model –as is common in the existing literature– the measured parameters may represent a reduced-form model simultaneously reflecting subjects’ IA and efficiency concerns. In contrast, our method facilitates measuring the IA parameters in both the IA model and the general model. A lack of robustness in the measurement of the IA parameters across the two models would then indicate that efficiency concerns are present.

Next, we introduce the production game that we will use to test the predictive power of the two models. This game is played by two players, Worker A and Worker B. At the start of the game, each receives a basic salary ( $s_i$ ,  $i = A, B$ ). 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  $e_i \in [0, e_{\max}]$ ,  $i = A, B$ . Department  $i$ ’s production  $p_i$  depends on the effort exerted by worker  $i$  in the following way:

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

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

$$\pi_i(e_A, e_B) = s_i + \frac{1}{2} \sum_{j=A,B} p_j(e_j) - e_i c_i, \quad i = A, B. \quad (5)$$

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

---

<sup>10</sup> Blanco et al. (2011) also note that the parameter describing a concern for efficiency is not identifiable if the environments used to measure are restricted to two-player games. Engelmann (2012) goes on to argue that a comparison of games with distinct numbers of players yields “rather implausible predictions” in the sense that a player will typically be inequality averse for some group sizes and inequality-loving for others. It goes too far to address this point in detail, here, but it suffices to say that we disagree that such an outcome is a priori implausible. For our purpose, this discussion is irrelevant, however, since we are solely interested in the robustness of the IA parameters.

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 Appendix A for details). Therefore, when applying model (1) to the production game, only the guilt parameter  $\beta$  for A and the envy parameter  $\alpha$  for B are relevant. When applying (3), the efficiency parameter is also relevant.

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

$$e_A = \begin{cases} 0 & \text{if } \beta_A \leq 0 \\ 200\beta_A & \text{if } 0 < \beta_A < \frac{1}{2} \\ 100 & \text{if } \beta_A \geq \frac{1}{2}. \end{cases} \quad (6)$$

Similarly, B has an optimal strategy in the production game:

$$e_B = \begin{cases} 0 & \text{if } \alpha_B \geq 1 \\ 100(1 - \alpha_B) & \text{if } 0 < \alpha_B < 1 \\ 100 & \text{if } \alpha_B \leq 0. \end{cases} \quad (7)$$

Therefore, if the IA model describes workers' preferences, A's effort choice in the production game is a dominant strategy completely determined by the guilt parameter  $\beta$ , while B has a dominant effort level solely dependent on her envy parameter  $\alpha$ .

The production game is particularly suited to test the inequity aversion model (1). There are at least four reasons why this is the case. First, as mentioned, both players have dominant strategies. Because their optimal choice requires no beliefs about the other worker's actions, attitudes towards risk and uncertainty are irrelevant for the model's predictions. This allows for a direct test of these predictions. Second, for any combination of effort choices, Worker A earns more than Worker B. As a consequence, each worker's predicted effort depends only on one kind of inequity aversion, again allowing for a direct test of the prediction. Third, the prediction is a continuous function of the relevant inequity aversion parameter (as opposed to the bang-bang corner predictions obtained for most games). This property facilitates a sharper test of the prediction. Finally, one can distinguish between a version of the production game where both players simultaneously decide on their effort choice and a sequential version where worker B sees A's choice before deciding herself. By

distinguishing between these two, one can isolate the effects of reciprocity. In the simultaneous game, the players cannot observe each other's action, and their decision is only dependent on their own preference levels, so reciprocity cannot play a role. In contrast, when Worker A chooses first, Worker B may condition her choice on A's chosen effort. The sequential game allows Worker B to respond to her perception of A's kindness, allowing for reciprocity to play a role in her decision.

The effect of introducing efficiency concerns into the model is also straightforwardly derived. If the workers' preferences are as in (3), the equilibrium effort levels are still in dominant strategies. More specifically, it can be derived (cf. Appendix B) that<sup>11</sup>

$$e_A = \begin{cases} 0 & \text{if } \beta_A \leq -\gamma_A \\ \frac{200(\beta_A + \gamma_A)}{1 + 2\gamma_A} & \text{if } -\gamma_A < \beta_A < \frac{1}{2} \\ 100 & \text{if } \beta_A \geq \frac{1}{2}, \end{cases} \quad (8)$$

and

$$e_B = \begin{cases} 0 & \text{if } \alpha_B \geq 1 + 3\gamma_B \\ \frac{100(1 - \alpha_B + 3\gamma_B)}{1 + 2\gamma_B} & \text{if } \gamma_B < \alpha_B < 1 + 3\gamma_B \\ 100 & \text{if } \alpha_B \leq \gamma_B. \end{cases} \quad (9)$$

Once again, envy does not enter A's optimal strategy and guilt does not affect B's. Note that the effort of both workers is increasing in  $\gamma$ : The more workers care about efficiency, the more effort they will exert. This is intuitive because the greater  $\gamma$  the more a worker will take into account that her effort increases both players' payoffs.

When discussing our results, we will use the production game to test the predictive power of model (1). We refrain from testing the prediction of model (3) as given in (8) and (9), because the aim of this paper is to test the robustness of the IA model.

### 3. Experimental Design

In the experiment, we use three choice menus to directly measure each subject's social preferences (i.e., to obtain estimates of  $\alpha$ ,  $\beta$ , and  $\gamma$ ). Subsequently, we test the

---

<sup>11</sup> The predictions for the efficiency model (2) follow by substituting  $\alpha = \beta = 0$ .

predictive power of the models for behavior in the production game. Transcripts of instructions are presented in Appendix C. We start by describing the menus used.

### Choice Menus

We first introduce the three choice menus used to measure social preferences.

#### Menu 1

Menu 1 is used to measure parameter  $\alpha$  of the inequity aversion model (1) and (together with Menus 2 and 3) the parameters of the general model (3). This menu consists of 10 decisions (cf. Table 1). 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 will be explained below). Each participant decides as if she is a proposer, because roles are not (randomly) determined until the end of the experiment.

**Table 1: Menu 1**

<b>Nr.</b>	<b>Option A</b>	<b>Option B</b>	<b>Choose B iff:</b>
1	Yours: 125 ; Other’s: 150	Yours: 100 ; Other’s: 260	$\alpha \leq -0.19$
2	Yours: 115 ; Other’s: 150	Yours: 100 ; Other’s: 260	$\alpha \leq -0.12$
3	Yours: 105 ; Other’s: 150	Yours: 100 ; Other’s: 260	$\alpha \leq -0.04$
4	Yours: 95 ; Other’s: 150	Yours: 100 ; Other’s: 260	$\alpha \leq 0.05$
5	Yours: 85 ; Other’s: 150	Yours: 100 ; Other’s: 260	$\alpha \leq 0.16$
6	Yours: 75 ; Other’s: 150	Yours: 100 ; Other’s: 260	$\alpha \leq 0.29$
7	Yours: 65 ; Other’s: 150	Yours: 100 ; Other’s: 260	$\alpha \leq 0.47$
8	Yours: 55 ; Other’s: 150	Yours: 100 ; Other’s: 260	$\alpha \leq 0.69$
9	Yours: 45 ; Other’s: 150	Yours: 100 ; Other’s: 260	$\alpha \leq 1.00$
10	Yours: 35 ; Other’s: 150	Yours: 100 ; Other’s: 260	$\alpha \leq 1.44$

*Notes.* The table presents the ten decisions (given in rows) between options A and B. “Yours” refers to the proposer’s payoff and “Other’s” to the receiver’s payoff. The final column includes the values of  $\alpha$  for which the inequity aversion model (1) rationalizes a choice of option A. Of course, this column was not shown to subjects.

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 1). From decision 4 onward, the proposer's own earnings are also lower in Option A than in B.

For decision 1, compared to Option B, disadvantageous inequality is 135 lower in A, while the proposer earns 25 more. Note that for non-negative  $\alpha$ , both remaining terms on the r.h.s. of (1) then imply higher utility for Option A than for B. In fact, any proposer with  $\alpha \geq -0.19$  will choose A.<sup>12</sup> 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 option A.

In this way model (1) determines for each decision question, a threshold for the proposer's envy level ( $\alpha$ ), 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 1. 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 Appendix D.

## Menu 2

Menu 2 also consists of 10 decision questions (cf. Table 2), 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  $\beta$  (cf. Appendix D). 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,

---

<sup>12</sup> Recall the assumption in the Fehr-Schmidt model that  $\alpha \geq 0$ . Choosing option B in decisions 1, 2, or 3 cannot be rationalized under this assumption.

**Table 2: Menu 2**

<b>Nr.</b>	<b>Option A</b>	<b>Option B</b>	<b>Choose B iff:</b>
1	Yours: 185 ; Other's: 90	Yours: 170 ; Other's: 50	$\beta \leq -0.60$
2	Yours: 175 ; Other's: 90	Yours: 170 ; Other's: 50	$\beta \leq -0.14$
3	Yours: 165 ; Other's: 90	Yours: 170 ; Other's: 50	$\beta \leq 0.11$
4	Yours: 155 ; Other's: 90	Yours: 170 ; Other's: 50	$\beta \leq 0.27$
5	Yours: 145 ; Other's: 90	Yours: 170 ; Other's: 50	$\beta \leq 0.38$
6	Yours: 135 ; Other's: 90	Yours: 170 ; Other's: 50	$\beta \leq 0.47$
7	Yours: 125 ; Other's: 90	Yours: 170 ; Other's: 50	$\beta \leq 0.53$
8	Yours: 115 ; Other's: 90	Yours: 170 ; Other's: 50	$\beta \leq 0.58$
9	Yours: 105 ; Other's: 90	Yours: 170 ; Other's: 50	$\beta \leq 0.62$
10	Yours: 95 ; Other's: 90	Yours: 170 ; Other's: 50	$\beta \leq 0.65$

*Notes.* The table presents the ten decisions (given in rows) between options A and B. “Yours” refers to the proposer’s payoff and “Other’s” to the receiver’s payoff. The final column includes the values of  $\beta$  for which the inequity aversion model (1) rationalizes a choice of option A. Of course, this column was not shown to subjects.

Option A gives the proposer more (185) and yields lower inequality (95) than B. Any non-negative  $\beta$  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 2 gives these threshold values for  $\beta$ .

### Menu 3

Together with Menus 1 and 2, Menu 3 (cf. Table 3) is used to measure the parameters of the general model (3). This menu, again, presents 10 decision questions to a proposer, distinguishing between Options A and B. In contrast to Menus 1 and 2, the proposer is now (anonymously) grouped with five receivers; each option specifies one payoff amount for the proposer and another amount for each of the receivers. This allows us to separate efficiency concerns from inequity aversion. Once again, each subject makes a decision as a proposer. At the end of the experiment, one group of six will be randomly selected and one of these six will be randomly selected as the proposer whose decision is implemented.

In each of the ten decisions, option A is the same, giving 50 to the proposer and to each of the five receivers (for aggregate earnings equal to 300). Compared to A, Option B gives less (25 in each decision) to the proposer. Moving down the menu, it gives increasingly more to each of the receivers, thereby increasing aggregate ear-

**Table 3: Menu 3**

Nr.	Option A	Option B	Choose B iff.
1	Yours: 50 ; Five Others': 50	Yours: 25 ; Five Others': 75	$\gamma \geq 0.250$
2	Yours: 50 ; Five Others': 50	Yours: 25 ; Five Others': 85	$\gamma \geq 0.167$
3	Yours: 50 ; Five Others': 50	Yours: 25 ; Five Others': 95	$\gamma \geq 0.125$
4	Yours: 50 ; Five Others': 50	Yours: 25 ; Five Others': 105	$\gamma \geq 0.100$
5	Yours: 50 ; Five Others': 50	Yours: 25 ; Five Others': 115	$\gamma \geq 0.083$
6	Yours: 50 ; Five Others': 50	Yours: 25 ; Five Others': 125	$\gamma \geq 0.071$
7	Yours: 50 ; Five Others': 50	Yours: 25 ; Five Others': 135	$\gamma \geq 0.062$
8	Yours: 50 ; Five Others': 50	Yours: 25 ; Five Others': 145	$\gamma \geq 0.056$
9	Yours: 50 ; Five Others': 50	Yours: 25 ; Five Others': 155	$\gamma \geq 0.050$
10	Yours: 50 ; Five Others': 50	Yours: 25 ; Five Others': 165	$\gamma \geq 0.045$

*Notes.* The table presents the ten decisions (given in rows) between options A and B. “Yours” refers to the proposer’s payoff and “Five Others’” to each of the five receiver’s payoffs. The final column includes the values of  $\gamma$  for which the efficiency model (2) rationalizes a choice of option A. Of course, this column was not shown to subjects.

nings. In decision 1, aggregate earnings (325) are only slightly higher in B than in A. Hence, a strong preference for efficiency is needed to give the higher utility to B. Note that from model (3) it follows that the threshold value of  $\gamma$  to switch from A to B depends on the inequity aversion measured by  $\alpha$  and  $\beta$ . Restricting social preferences to only efficiency concerns (model (2)) allows one to directly determine a lower bound for  $\gamma$  for each decision, however. These bounds are shown in the last column in Table 3. Our main interest lies in testing the robustness of inequity aversion estimates to allowing for efficiency concerns (i.e., model (3)), however, and not in the efficiency model *per se*.

### Combining Menus 1-3

The above is based upon the assumption that subjects have either only envy and guilt (Menu 1 and Menu 2) or only efficiency (Menu 3) concerns. In particular, the  $\alpha$ - and  $\beta$ - [ $\gamma$ -] cutoff points in Tables 1-3 are based on the partial model (1) [(2)]. If we consider the general model (3), efficiency concerns may play a role when choosing in Menus 1 and 2 and inequity aversion may influence the choices in Menu 3. By combining an individual’s choices in all three menus, we can jointly estimate  $\alpha$ ,  $\beta$ , and  $\gamma$ . Moreover, comparing subjects’ choices in the three menus will allow us to evaluate the robustness of the Fehr-Schmidt model to the inclusion of efficiency concerns. More in particular, we will estimate the extent to which estimates obtained with the

Fehr-Schmidt model alone (model 1) are consistent with measures obtained in the general model. Details on this procedure are given in Appendix D.

### Production Game

When introducing the production game to subjects, equations (4) and (5) were (of course) not used. Instead, participants were 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. We use the strategy method: when deciding, the subjects do not know which role they will have, and 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.

**Diagram 1: Production Game Calculator**

**Part 4: Production Game**

**Worker A's Decision**

Suppose you are Worker A. Your effort level will influence the payoffs of you and your paired participant (Worker B).

Draw the scrollbar to choose your decision on Worker A's effort level.

In the table below the scrollbar, the amount of Bonus 1 and the effort cost due to your chosen effort level is displayed.

**Worker A's Effort Level** 0  100

Worker A's Effort:	50
Bonus 1 for each worker:	88
Effort cost for Worker A:	100
Basic Salary+Bonus 1-Effort cost for Worker A:	188
Basic Salary+Bonus 1 for Worker B:	88

**Note that earnings may also be affected by Worker B's decision.**

**Worker B's Decision**

Suppose you are Worker B. Your effort level will influence the payoffs of you and your paired participant (Worker A).

Draw the scrollbar to choose your decision on Worker B's effort level.

In the table below the scrollbar, the amount of Bonus 2 and the effort cost due to your chosen effort level is displayed.

**Worker B's Effort Level** 0  100

Worker B's Effort:	29
Bonus 2 for each worker:	54
Effort cost for Worker B:	29
Basic salary+Bonus 2 for Worker A:	254
Basic salary+Bonus 2-Effort cost for Worker B:	25

**Note that earnings may also be affected by Worker A's decision.**

Press **OK** to submit your final decisions on the effort levels of Worker A and Worker B.

OK

For each role, the participant can try out any effort levels by moving a scroll bar. The table directly 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 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.

We used two versions of the production game, which we varied across subjects. In the first, subjects were not informed about their roles, and made their decisions simultaneously for both Worker A and B. We denote this simultaneous production game by SimProd. In contrast, subjects make decisions in the production game sequentially in our treatment SeqProd. For this, we again use a strategy method. 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 Appendix C show how participants were informed about using this 'Decision Box'.

**Diagram 2: Decision Box for B's Conditional Effort**

**Decision Box**

**Rule Creator**

Upper Bound

Effort

Add into the rule list

**Responding Rules**

Lower Bound	Upper Bound	Effort
0	8	1
9	50	13
51	100	73

Delete

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, which will be only revealed at the end of the experiment, will be implemented.<sup>13</sup>

### **Experimental Procedures**

The experiments were conducted at the CREED laboratory of the University of Amsterdam (UvA). Subjects were recruited from the CREED subject pool, which consists of approximately 2000 students, mainly UvA 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. Parts 1, 2, and 3 measure social preferences using Menus 1, 2, and 3 respectively. In Part 4, the production game is played. Subjects do not learn about their roles and their payoffs in any part until the end of the experiment.

In Part 1, every subject is asked to make her choices for Menu 3. Groups of six are randomly formed.<sup>14</sup> Subjects know that one of the ten decision questions will be randomly selected and that one of the six subjects in each group will be randomly assigned to be the proposer. This proposer's decision for the selected question is implemented for her group. Subjects are also told that if the number of subjects is not a multiple of six, some will not be allocated to a group, and hence are not paid for this part. Of course, when making their decision, they do not know whether or not they have been allocated to a group.

In Part 2, subjects are randomly assigned into pairs. Each subject is asked to make choices for Menus 1 and 2. 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 across subjects), different payoffs are implemented. Specifically, the payoffs of Menus 1 and 2 are multiplied by 10, 30 and 60, compared to part 2. In what

---

<sup>13</sup> As will become clear, we used these variations of the production game (simultaneous versus sequential and single versus double roles) to isolate the effect of reciprocal concerns on inequity aversion.

<sup>14</sup> This randomization was done before the experiment took place.

follows, we denote these treatments by x10, x30 and x60, respectively. 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 this part, 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.<sup>15</sup>

Our focus in Parts 1-3 of the experiments is on the social preference models (1)-(3). Whereas choices in Parts 1 and 2 allow us to obtain parameter estimates for these models in the way described above, the scaling of payoffs in Part 3 allow us to test the underlying linearity assumptions. For example, in the Fehr-Schmidt (1999) inequity aversion model, the marginal disutility of inequality is independent of the payoff level. This means that measures of  $\alpha$  and  $\beta$  obtained from Parts 2 and 3 should be equal.

Finally, in Part 4, we randomly paired subjects and let them play the production game described above, either as SimProd or as SeqProd. As described above, this will be used to test the predictive power of the estimates. For each of the payoff levels of Part 3, we ran a low payoff simultaneous production game with payoffs in the order of magnitude of Parts 1 and 2 and a high payoff version with payoffs in the magnitude of Part 3 in the session concerned. We denote the former by p1 and the latter by p10, p30, or p60. These were varied across subjects.

Table 4 summarizes our treatment combinations and shows how many observations we have for each cell. 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.

Aside from the treatments depicted in Table 4, we also have two control treatments. One, denoted by MenuOnly containing only Part 3 (x10), was used to check whether the experience in the low-scale parts 1 and 2 has influence on subjects' decisions in the high-scale menu tests. The other, ProdOnly, consisted of only the (simultaneous) production game of part 4 (p30). This treatment was used to check

---

<sup>15</sup> As explained below, 50 subjects participated in control treatments where this decision did not need to be made. In our discussion of the results, we will distinguish 179 subjects whose decisions were 'rationalizable' in various ways. All of these subjects chose to participate in part 3.

**Table 4: Treatments**

Part 3:	Part 4:	
	SimProd	SeqProd (payoff p10)
x10	p1 (28), p10 (34)	DoubleRole (30), SingleRole (30)
x30	p1 (30), p30 (30)	---
x60	p1 (30), p60 (22)	---

*Notes.* Numbers in parentheses indicate the number of (independent) observations. In all treatments, Parts 1 and 2 took place as described in the main text. The multiplication factor used in Part 3 is denoted by the treatment variable in the rows and in Part 4 the production game is played either simultaneously or sequentially, as indicated by the columns. Payoffs in the simultaneous production game are scaled either as x1 or as the corresponding scale in part 3. All treatments presented in the table were varied across subjects.

whether having participated in the menu tests or not influences subjects' behavior in the production game.<sup>16</sup>

#### 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 allowance of efficiency considerations (model 3) by including Menu 3 and to the scaling of payoffs.<sup>17</sup> 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 estimates to reciprocity concerns.

#### Estimates of Envy and Guilt

Our first estimates of the envy parameter ( $\alpha$ ) are derived from Menu 1. Specifically, the threshold values shown in the last column of Table 1 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

<sup>16</sup> Of the 284 participants, 24 participated in MenuOnly and 26 in ProdOnly. Average earnings were 12.45 in MenuOnly, 38.79 in ProdOnly as compared to 41.30 across all other treatments. We do not find any evidence indicating that experience in the standard-scaled menu tests (Part 1 and 2) affects decisions in the high-scaled 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 0.99 for Menu 1 and 0.95 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 0.25 and 0.67 for Worker A's and B's effort levels, respectively). We will therefore not further discuss these control sessions.

<sup>17</sup> Because preferences for efficiency *per se* are not the main interest of this paper, estimates of  $\gamma$  are presented in Appendix E, which derives such estimates using either the efficiency model (2) or the general model (3). We basically find that little importance is attributed to efficiency.

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”.<sup>18</sup> Amongst the group of 216 IA-consistent subjects, 103 are estimated to have a negative value of  $\alpha$  either in the standard-scale or the high-scale environment, which would indicate a preference for increased (disadvantageous) inequity aversion. Note that this possibility is excluded by assumption in the Fehr-Schmidt model. However, as we will show later, 66 of these 103 subjects can be rationalized to have non-negative envy level by using a model that also allows for efficiency concerns. The remaining 37 subjects are denoted as “non-IA-rationalizable” since their choices are inconsistent with the inequity aversion framework.<sup>19</sup> We will exclude the 18 IA-inconsistent and the 37 non-IA-rationalizable observations from the further data analysis, leaving 179 IA-consistent-and-rationalizable (IA-C&R henceforth) observations.<sup>20</sup> For completeness’ sake an analysis applied to the set of 216 IA consistent subjects (and thus including the non-IA-rationalizable observations) is provided in Appendix F. Our main conclusions turn out to be unaffected by this choice.

On the other hand, we have decided not to exclude subjects with  $\beta$  estimates that contradict the restriction  $\beta \in [0, \alpha] \cap [0, 1)$  proposed by Fehr and Schmidt (1999). In fact, many subjects violate this condition. 65 participants (36.3% of the IA-C&R subjects) have an estimated  $\beta$  level greater than their  $\alpha$ .<sup>21</sup> We see no theoretical reason for this assumed restriction, however, and the IA model can easily be applied without it. We therefore include these subjects in our analyses. Finally, one subject has a

---

<sup>18</sup> 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.

<sup>19</sup> 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 a partial inequity aversion model like (1), such a subject may be perceived as having a negative envy level. For this reason, we relax the original restriction on envy ( $\alpha \geq 0$ ) and only enforce it for the general model (3). In other words, non-IA-rationalizability refers to negative envy, even after correcting for efficiency concerns.

<sup>20</sup> Recall that each of these subjects chose to enter part 3 of the experiment.

<sup>21</sup> This is very close to the result reported by Blanco et al. (2011), that 23 (37.8%) of their 61 subjects violate the assumption that  $\alpha \geq \beta$ .

negative estimated  $\beta$  level, which is estimated to lie in the range  $(-0.60, -0.14]$ .<sup>22</sup> A reason to include this observation is that we think it in fact reasonable to exhibit (slightly) negative guilt levels, a preference indicating that being slightly better off increases one’s utility. In contrast, we see no such ‘reasonable interpretation’ for negative envy levels.

Our estimates of the IA-C&R 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).<sup>23</sup> Our results differ substantially from those previously found.  $\chi^2$ -tests reject the null-hypotheses that the distribution of our estimates for  $\alpha$  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 (97.8%) 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.

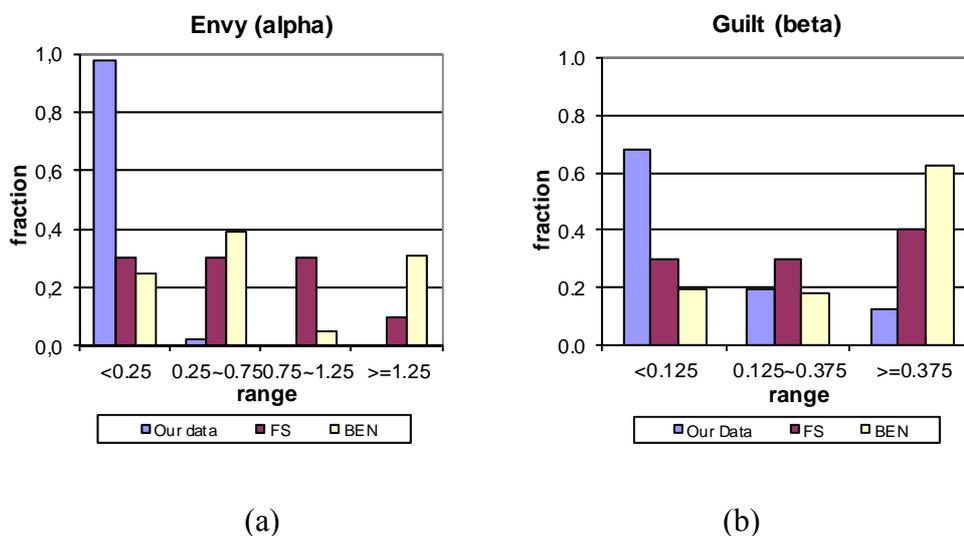


Figure 1. Distribution of Envy (a) and Guilt (b) of IA-C&R subjects estimated according to IA model by us, FS and BEN

<sup>22</sup> In addition, 122 subjects are estimated to have a  $\beta$  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 2, 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.

<sup>23</sup> The distribution given by FS has a few points with mass density (in particular,  $\alpha=0, 0.5, 1$ , and  $4$ ). For comparison, we summarize our estimates in intervals of which the midpoints coincide with the FS estimates (except for the highest interval  $[1.25, +\infty)$  which contains the FS estimate of  $4$ ). The intervals reported by BEN are different; therefore we recategorized their observations using the information provided in their online appendix.

In a similar way, we use Menu 2 to provide a first estimate of the guilt parameter,  $\beta$ . Our estimates are presented and compared to the FS and BEN estimates in Figure 1(b).<sup>24</sup> 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 towards the left. The null-hypotheses of the distribution of our  $\beta$  estimates being the same as the distributions estimated by FS and BEN, are both rejected at the 0.01 level.

Before testing the predictive power of our estimates, we first test their robustness to scale and efficiency concerns.

### Robustness to Scale and Efficiency concerns

Because all of the IA-C&R subjects chose to enter Part 3, we have 179 observations for which we can compare  $\alpha$  and  $\beta$  estimated under different payoff scales. To start, Figure 2 shows the estimated distributions for the various payoff scales used. Recall that this is a within-subject comparison: each subject participated in the benchmark scale (x1) and in one of the higher scale treatments.

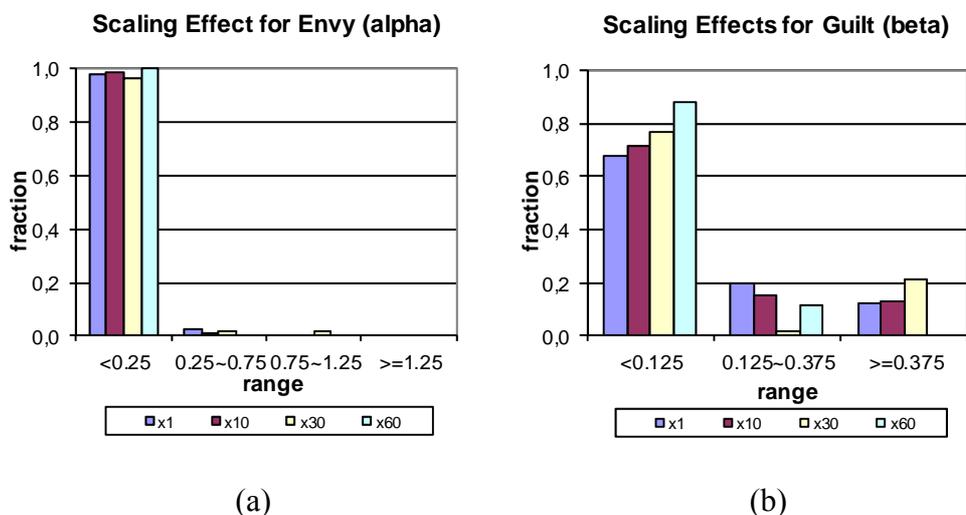


Figure 2. Distributions of Envy Parameters (a) and Guilt Parameters (b) of IA-C&R subjects for distinct payoff scales

A first impression from the figure is that estimates of the envy parameter are insensitive to changes in the payoff scale, though this may be due to the almost complete lack of envy in the first place. For guilt, there appears to be a shift towards lower  $\beta$ -values with increasing scale. To test whether this effect is statistically

<sup>24</sup> Again, we center our intervals around the FS estimates, in this case  $\beta=0, 0.25, \text{ and } 0.6$ .

relevant, we use Wilcoxon sign-rank tests comparing at the individual level the IA parameters obtained from the high and low scale menus. The results are summarized in Table 5.

**Table 5: Scale effects**

	x10 (84 obs.)		x30 (52 obs.)		x60 (43 obs.)	
	z	p-Value	Z	p-Value	z	p-Value
$\alpha$	0.527	0.598	0.495	0.620	1.613	0.107
$\beta$	0.543	0.587	-1.463	0.144	-3.155	0.002***

*Notes.* Results are presented from Wilcoxon Sign-rank tests of  $H_0: \alpha_{xi} = \alpha_{li}$  vs.  $H_1: \alpha_{xi} \neq \alpha_{li}$ , and  $H_0: \beta_{xi} = \beta_{li}$  vs.  $H_1: \beta_{xi} \neq \beta_{li}$ ,  $x \in \{10, 30, 60\}$ .  
\*\*\*indicates statistical significance at the 0.01 level.

For the estimates of the envy parameter, the test results do not reject the null hypothesis that they are invariant to changes in scale. Therefore, we do not reject the hypothesis that the disutility from the payoff difference is linear in disadvantageous inequality. For the estimates of the guilt parameter, the test also does not reject the robustness of the estimates when payoffs are scaled up by a factor 10 or 30. We do, however reject the null of no scale effect when payoffs are 60 times higher than in the benchmark case. 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. The linearity assumption for advantageous inequity is warranted for moderate increases in payoff levels, but not for major 60-fold increases.

Next, we consider the robustness of our results to allowing for efficiency concerns. We do so by deriving estimates of  $\alpha$  and  $\beta$  from the general model (3) (as explained in Appendix D) and comparing these to the estimates from the standard model (1). Figure 3 compares the derived distributions of  $\alpha$  and  $\beta$ . At first sight, the introduction of efficiency concerns does not affect the distribution of the envy parameters. For the guilt parameter, the General model estimates put slightly more weight in the lowest range than the partial model estimates. The difference is not significant, however. According to the given categories as shown in the figures,  $\chi^2$  tests give a  $p$ -value of 0.27 for  $\alpha$  and 0.68 for  $\beta$ , so we cannot reject the null hypothesis that the distributions of the IA estimates are the same as the General model estimates<sup>25</sup>.

<sup>25</sup> If we apply a one-to-one comparison of estimated parameters from the two models at the individual level, a signed-rank test rejects the null hypothesis that the parameters are equal for both  $\alpha$  and  $\beta$ .

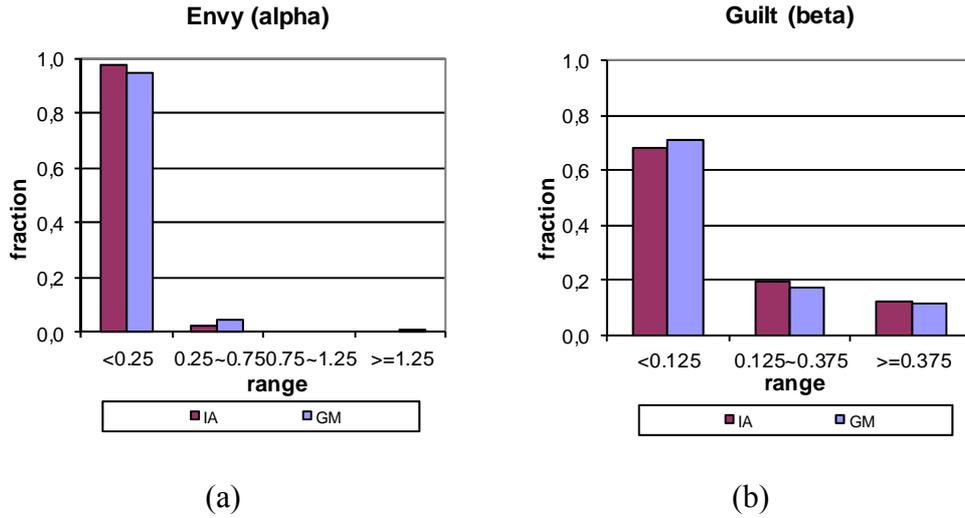


Figure 3: Distribution of Envy ( $\alpha$ ) and Guilt ( $\beta$ ) parameters of IA-C&R subjects after correcting for efficiency concerns

As explained in Appendix D, a direct comparison of point estimates for  $\alpha$  and  $\beta$  derived using the two models (1) and (3) has drawbacks. For example, a point estimate for the IA model (1) is derived from the intervals for  $\alpha$  implied by the last columns in Tables 1 and 2. These intervals change if we allow for efficiency concerns in the general model (2). A particular switching point in Menu 1 then yields a (slightly) different interval for  $\alpha$ . An alternative way to check the robustness of the IA estimates to the allowance of efficiency concerns is to use the *Overlap Ratio* (see Appendix D). This ratio, denoted by  $\Omega$ , measures the average overlap of the intervals measured using models (1) and (3) as a percentage of the interval measured by the IA model (1).<sup>26</sup>  $\Omega = x\%$  then indicates that on average  $x\%$  of the estimated  $\alpha$  (or  $\beta$ ) 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 (or 2) and 3. Weighted by the number of observations, we find  $\Omega = 82.1\%$  for  $\alpha$  and  $\Omega = 88.5\%$  for  $\beta$ .

An alternative measure for the robustness of IA parameter estimates to allowing for efficiency concerns simply determines the fraction of cases for which the interval derived from the partial IA model overlaps with the interval derived from the general

---

However, this is strongly affected by a methodological artifact, i.e., the test depends strongly on the point estimates chosen for both models, as explained below. Any method used to derive a point estimate from an interval will yield distinct parameter estimates for the two models, which strongly influences the test results. We therefore do not apply this kind of one-by-one comparison. As a robustness check for the tests discussed in the main text, the next two paragraphs introduce alternative ways to test the consistency between the estimates generated by the two models.

<sup>26</sup> Because multiple observations exist for various combinations of switching points in Menus 1 (or 2) and 3, we weigh observed overlap by the numbers of observations.

model (as described in Appendix D). An overlap means that there are values of  $\alpha$  ( $\beta$ ) that allow a subject's choices to be rationalized by both models. 88.8% of the IA-C&R subjects fulfill this criterion (cf. Appendix D.III).

All in all, our estimates of IA parameters are quite robust to allowing for efficiency concerns. This can partly be attributed to the finding that our subjects show less concern for efficiency (see Appendix E) than observed in some previous studies (e.g., Charness and Rabin 2002; Engelmann and Strobel 2004). A possible explanation for this difference with previous studies is that we are the first to directly measure preferences for efficiency using simple individual choice menus.<sup>27</sup> We will not further dwell upon possible reasons for these differences, because our main interest is in the robustness of the IA model and not in efficiency concerns *per se*.

### **Predictive Power of IA model in the Production Game**

We can use the estimates of an individual's  $\alpha$  and  $\beta$  to predict her behavior in the production game. Recall that we have two versions of the production game, SimProd and SeqProd (cf. Table 4). To test whether the predictions for the production game are supported by the observations from the lab, we ran the following regression:

$$e_{R,i}^T = \delta_{R,0}^T + \delta_{R,1}^T \alpha_{x1,i}^{IA} + \delta_{R,2}^T \beta_{x1,i}^{IA} + \varepsilon_{R,i}^T \quad (10)$$

where  $T = \text{SimProd}, \text{SeqProd}$ ,  $R = A, B$  represents the subject's role in the production game,<sup>28</sup> and  $i$  is the index of a subject.  $\alpha_{x1,i}^{IA}$  and  $\beta_{x1,i}^{IA}$  are the IA estimates of  $i$ 's envy and guilt parameter (as derived from Menus 1 and 2), respectively. Table 7 gives the estimated coefficients of the  $\delta$ 's in (10).

Comparing the results in Table 6 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 only 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. For

---

<sup>27</sup> Charness and Rabin (2002) use a series of dictator choices to test subjects' preferences, but without trying to measure the degree of preference for efficiency (for which they use the term "social welfare").

<sup>28</sup> Recall that in most treatments, subjects played both roles.

**Table 6: Coefficients of IA parameters in the production game.**

	IA Prediction		SimProd		SeqProd	
	$e_A$	$e_B$	$e_A$ (obs. 144)	$e_B$ (obs. 146)	$e_A$ (obs.14)	$e_B$ (obs. 21)
const.	0	100	29.70*** ( 3.24)	95.84*** ( 1.36)	68.31*** (8.61)	76.80*** (8.05)
$\alpha$	0	-100	-4.58 (31.85)	-61.28*** (13.35)	146.98 (147.11)	-245.91 (157.10)
$\beta$	200	0	100.52*** ( 15.44)	-9.91 ( 6.47)	-1.60 (31.60)	-140.98** ( 62.11)

*Notes:* The predicted coefficients are taken from eqs. (6) and (7). Numbers in parentheses give standard errors. \*\*\* indicates statistical significance at the 1%-level.

the two treatments of the sequential production game, the regression outcome differs substantially from the theoretical prediction, however. We find either that none of the predicted effects exists (for Worker A's effort), or that the significance is exactly opposite to the theoretical prediction (only the guilt levels have significant impact, for Worker B's effort).<sup>29</sup> We will return to this difference between the two games when discussing our results below.

Using the chosen effort levels, we can also obtain a direct estimate of individuals' envy ( $\alpha$ ) and guilt ( $\beta$ ) parameters by using the inverse relationships of (6) and (7).<sup>30</sup> The resulting distributions are displayed in Figure 4. 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 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 (21 observations in total) are used to estimate  $\alpha$ ; similarly, only the realized effort levels by workers A (14 observations in total) are used to estimate  $\beta$ .

<sup>29</sup> For the regression in SeqProd, for Worker B's effort we use the realized levels, according to their responding rules in response to their paired Worker A's effort levels. Also, we use only the true Worker A's effort chosen for A.

<sup>30</sup> 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.

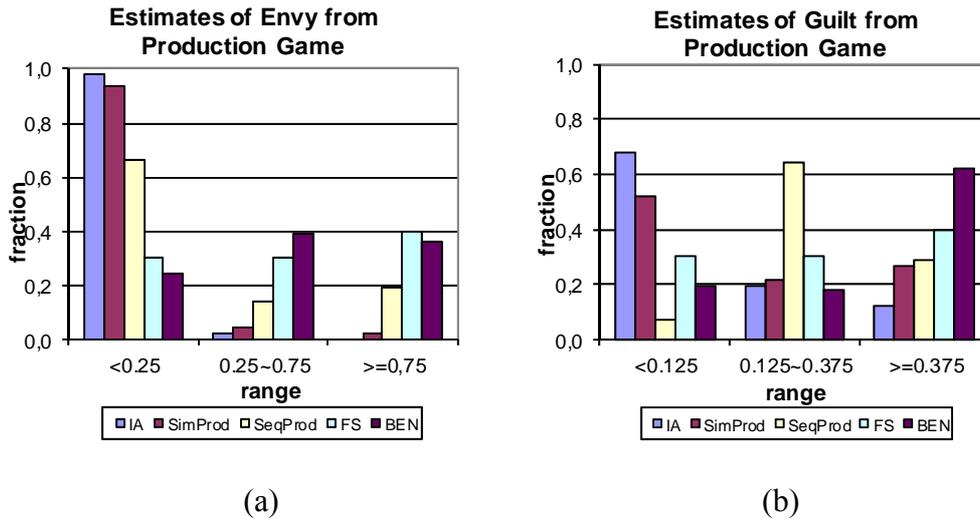


Figure 4: Distribution of the envy (a) and guilt (b) of IA-C&R subjects derived from the production game.

Figure 4(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 the reciprocity-free environment of SimProd, the envy parameters are very similar to those estimated from Menu 1 (which is reciprocity free in a trivial sense). This explains why the regression results reported for SimProd in Table 3 are quite consistent with the theoretical prediction.

The results are different for the sequential production game, where reciprocity is possible. Here, far fewer subjects (66.7% as opposed to 93.8% 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 seems to imply 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 that their envy parameters were also derived from an environment where subjects may be exposed to negative reciprocity.

In Figure 4(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

Two things stand out in our results. First, our estimates of disadvantageous inequity aversion (envy) and advantageous inequity aversion (guilt) are robust to allowing for efficiency concerns and also reasonably robust to increases in the scale of payoffs (though guilt seems to become 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. When we introduce the possibility to reciprocate, we observe more subjects for whom envy plays a role, however.

This effect of reciprocation in our production game shows up in changes in both workers' behavior in the sequential game compared to the simultaneous one. 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, which 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  $\alpha_i$ 's and  $\beta_i$ '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., with the possibility to reciprocate) 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 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.

All in all, we believe that the control for 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 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 reciprocity is (more or less) absent. However, when reciprocity is involved, 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 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.

## References

- Binmore, K. and Shaked, A. 2010. "Experimental economics: Where next?" *Journal of Economic Behavior & Organization*, 73 (1): 87-100.
- Blanco, M., Engelmann, D. and Normann, H.T. 2011. "A within-subject analysis of other-regarding preferences". *Games and Economic Behavior*, 72 (2): 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. 2011. "The impact of relative position and returns on sacrifice and reciprocity: An experimental study using individual decisions". Working paper.
- Camerer, C. 2003. "Behavioral game theory: Experiments in strategic interaction". Princeton University Press, Princeton.
- 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. 2009. "Other regarding preferences: a selective survey of experimental results". Forthcoming in the *Handbook of Experimental Economics*. 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. 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.

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.