EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

EUI Working Paper **ECO** No. 2005 /11

Fairness vs. Social Welfare in Experimental Decisions

STEFAN KOHLER

BADIA FIESOLANA, SAN DOMENICO (FI)

Published in Italy in October 2005
European University Institute
Badia Fiesolana
I-50016 San Domenico (FI)
Italy

Fairness vs. Social Welfare in Experimental Decisions*

Stefan Kohler†

First version: June 7, 2002
This version: September 22, 2005

Abstract

Experimental evidence from modified dictator games and simple choice situations indicates that concern for overall welfare is an important motive in human decision making. Models of inequality averse agents, as suggested by Fehr and Schmidt (1999) or Bolton and Ockenfels (2000), fall short in explaining behavior of proposers, who reduce their payoff below a fair split of the endowment to maximize social welfare, while other types of social preferences do well on these data. This has created the impression that inequality aversion is a misguided concept. This paper presents a formal model and shows that a combination of welfare concern and inequality aversion changes this result in favor of inequality aversion. It also establishes a unique link between altruism and social welfare in the proposed model.

JEL classification: A13, B49, C70, D63, D64
Keywords: Social Preferences, Inequality Aversion, Welfare Concern, Reciprocity

*I would like to thank Karl Schlag, Dorothea Herreiner, Eyal Winter and Stefan Imof for helpful comments. The paper benefitted also from discussions at the EUI and at the first Budapest Workshop on Behavioral Economics. All remaining errors are my own.

†European University Institute, Department of Economics, Villa San Paolo, Via della Piazzuola 43, 50133 Florence. E-mail: stefan.kohler@iue.it.

1 Introduction

Departures from purely self-interested behavior in economic experiments have inspired models of *social preferences*, where researchers experiment with different ways of analytically decoding individuals' concern for others. The main focus of this work is on decomposing experimentally observed concern for the outcomes of others into underlying primary behavioral motives. Social preferences formally depict other-regarding motivations of human decision making in addition to pure self-interest. Thereby they contrast with the canonical neoclassical economic models that are based on the hypothesis that all people are exclusively concerned about their own material well-being.[1]

At present, models of altruism, inequality aversion and reciprocity provide exclusive explanations of non-selfish behavior in some economic environments, but compete for explanatory power in others. While, for example, altruism, inequality aversion or reciprocity are a rational for sharing of a surplus as observed in Ultimatum Game experiments, only inequality aversion or reciprocity are consistent with the rejection of low offers. On the other hand, only inequality aversion and altruism provide explanations for unilaterally motivated transfers in Dictator Games. The intuition therefore is that inequality averse agents mind the initial unequal payoff allocation and altruists enjoy their own as well as others well-being. As the Dictator Game is a one-sided one-off decision without any interaction the intention-based concept of reciprocity does not apply to it. In turn, reciprocity is the only concept incorporating a progression of action. Inequality aversion and altruism, two outcome-based concepts, disregard this potentially important information. Thus, while an ad hoc success of all three of these social preference motives to explain experimental data in exclusive domains has manifested the validity of each concept as an individual behavioral motive, there is still much disagreement about the validity of each single motive whenever a strategic situation allows more concepts to be successfully applied.

A turning point in the discussion about the respective validity of inequality aversion and reciprocity seemed to be reached when Charness and Rabin (2002) confronted models of distributional preferences and reciprocity with data from a large scale experiment on simple decision problems between two payoff allocations. They found that overall-welfare was frequently maximized, but that this concern, including the other player's wealth, significantly decreased when a previous action was perceived as unkind. Both reciprocity and inequality aversion as exclusive social motives were inconsistent with a large fraction of the

[1] Self-interest throughout this paper strictly refers to this narrow interpretetion.

experimental choices they observed. Charness and Rabin interpret this as evidence against the Fehr and Schmidt (1999) model of pure inequality aversion and show conclusively that most of the observed behavior can be explained by reciprocity in combination with direct welfare concern. Combining reciprocity and welfare concern in one model, they report an overall consistency with the observed choices of 94 (95) percent of their approach, while the consistency of inequality aversion remains at 82 (86) percent. Still, both do significantly better than the neoclassical pure self-interest hypothesis with 73 (84) percent.[2] However, this evaluation of these two competing behavioral motives is biased in favor of their own model. With evidence at hand that welfare concern is a motive of its own, and their experimental games in which the welfare motive matters, a comparison should be done between two *hybrid models*.

The following paper suggests an inequality-averse social welfare maximizer (henceforth IASW) model and shows that it is an adequate counterpart. It fits well in a recent literature that has experimented with hybrid models[3] and demonstrates once the Fehr and Schmidt (1999) benchmark model is extended by welfare concern it is applicable to broader sets of experimental data, like the Charness and Rabin (2002) dataset, where a model of pure inequality aversion was misguided. The reason therefore is that inequality aversion on its own fails to express the tension individuals seem to feel, when social welfare can be maximized but difference thereby also increases. The IASW model offers insights in the interplay between these concerns, that can offset or reinforce each other. I also show that combining inequality aversion and welfare concern, Charness and Rabin (2002)'s consistency measure changes in favor of the IASW model (98 percent). Finally, the step of formalizing a model of how inequality aversion and welfare concern may interact induces conjectures that can be confronted with (new) experimental data. If the model is accepted it offers additional insights into the nature of preferences.

The paper is organized as follows: Section 2 presents experimental evidence for welfare concern. Section 3 reviews some recent social preference models. It closes with a formal depiction of altruism. Then section 4 suggests a new theoretical model (IASW) of inequality-aversion and social welfare concern. The model is comprehensively introduced on the Ultimatum Game and then confronted with further experimental data in section 5. Section 6 concludes.

[2] Different values result from different assumptions about first the players' knowledge.

[3] Charness and Rabin (2002) combine reciprocity and welfare concern, Falk and Fischbacher (2005) reciprocity and inequality aversion in a formal model.

2 Experimental Evidence of Welfare Concern

The simplest game to elicit welfare concern is a generalized version of the Dictator Game. In the standard version of the Dictator Game (henceforth DG) two players are allotted an initial sum of money, which for simplicity is normed to 1. The first player, called the *proposer* (or dictator), can unilaterally decide how to split this endowment by choosing a transfer t. The second player, called the *receiver*, then receives the share $t \in [0, 1]$ of the initial endowment that is allocated to him as determined by the dictator. In the generalized version of the Dictator Game the setting is very similar. Again the initial endowment to be divided by the dictator equals one. However, now the share which gets allocated to the receiver is multiplied by the experimenter at a previously announced exchange rate $1 : m$. Thus, in the generalized DG, final welfare is directly at stake through a connection to the allocative decision of the dictator.

figure 1 about here

Inequality aversion and altruism are the only motives that can explain an incentive to share at all in the DG. Concern for total welfare contains an altruistic component. It can be separated from inequality aversion by comparing transfers t in a generalized version of the DG at a varying exchange rate. If mean transfers increase disproportionate to welfare in m then the prospect of a higher overall welfare itself at higher m has triggered the increased transfer. This evidence for inequality aversion as at most a partial motive is particularly strong if also offers are made such that the inequality between players (again) increases; e.g. if shares that transfer more than half of the final total welfare are offered.

Empirically affirmative evidence for welfare concern as an independent behavioral motive was identified in generalized DG experiments with exchange rates varying from 1:1 to 1:4 by Andreoni and Miller (2002), Bolle and Kritikos (2001), Cox (2000) and Güth, Kliemt and Ockenfels (2002). The latter group of authors excepted, all these papers indicate a non-negligible fraction of players who were willing to give up a proportion of their own monetary payoff if they could increase total welfare. This proportion was the higher the more players could increase total welfare, i.e. $\partial t/\partial m > 0$. Such behavior was observed even when doing so implied that proposers generate inequality to their own disadvantage. In the standard DG where the exchange rate is 1:1, there are virtually no subjects who transferred more than 50 percent of the total welfare.[4] In contrast, when the exchange rate increases to 1:3, i.e. the experimenter triples the proposed transfer, Cox (2000), for instance, finds that a fraction of 60 percent make welfare-maximizing transfers such that they end up with less money than the receiver. Because disadvantageous inequality must be compensated for inequality averse players, this is strong evidence for independent welfare concern.

Other empirical evidence of welfare concern, also establishing a discrimination between the competing approaches of distributional and reciprocal preferences, is found in Charness and Rabin (2002, henceforth CR). CR confront models of distributional preferences and reciprocity with data from a large scale experiment on simple decision problems between two payoff allocations. In one and two-stage games they find that reciprocity and inequality aversion on their own as motives of social preferences were inconsistent with a large fraction of the experimental choices. In their experiment, in which the one-stage decision problems can be interpreted as generalized DGs with only two payoff situations, welfare at stake was frequently maximized even when the allocator had to trade off a proportion of his own payoff and doing so increased inequality; i.e. when welfare maximization was costly. When experimentees had to choose between (400,400) and (375,750), on average 66 percent chose the latter, welfare maximizing allocation. These choices are inconsistent with pure inequality aversion models since the difference between players increased, while the own payoff decreased. Neither they are consistent with any model of pure reciprocity, since a one sided one-off choice leaves no scope for reciprocal concerns. Again these choices reveal a direct and independent concern for welfare.

3 Review of recent Social Preference Models

Within the formal rational-choice models that explain the various experimental evidence two main approaches can be distinguished. The first focuses on *distributional preferences*, which means there are at least some people who care not only about absolute, but also about relative well-being. Distributional preferences assume that agents prefer to minimize differences between their own monetary payoff and those of other people. The second approach in contrast assumes *reciprocal preferences,* i.e. that agents care about the intentions behind an action, and that they want to reward kind and punish unkind behavior. This

[4] This is a reason why inequality aversion has more appeal than pure altruim as a behavioral motive.

section reviews recent models of social preferences and closes by introducing a formal model of altruism.[5]

3.1 Distributional Preferences

Models of distributional preferences have been proposed by Loewenstein, Bazerman, and Thompson (1989), Bolton (1991), Bolton and Ockenfels (2000), and Fehr and Schmidt (1999, henceforth FS). The last two approaches thereby offer the most readily applicable models. While FS model the intuition that individuals compare themselves to single members of a reference group, Bolton and Ockenfels (2000) believe that inequality averse agents compare themselves to a group average. Engelmann and Strobel (2004) test these competing hypotheses and find that experimental evidence is often in favor of the FS intuition.[6] In FS' model, which is the basis for the following inequality aversion and social welfare model, player $i \in [1, ..., n]$ has preferences of the following form:

$$u_i(x) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_i - x_j, 0\}. \quad (1)$$

There are n players and FS assume $\beta_i \leq \alpha_i$, $0 \leq \beta_i < 1$, where α_i and β_i are the parameters of the model measuring how much player i dislikes having less money or more money than others. This means that receiving a smaller payoff than others affects a person at least as much as getting more, and that he is never so affected by getting more than others as to want to throw away his own money without benefiting others. With that model FS can match experimental data in the Ultimatum Game, Public Good Games, Market Games (henceforth UG, PGGs, MGs) and some other games, and they can explain why there is an incentive for sharing in the DG.

3.2 Reciprocity Preferences

Reciprocity preferences motivated by the observation that players care about the intentions of other players, and that they are willing to sacrifice money to reward kind and to punish

[5] Early (partly formal) ideas of social preferences date back to Veblen (1934), Duesenberry (1949), Leibenstein (1950), Pollak (1976) and somewhat later the influential Rabin (1993).

[6] Furthermore CR find that more than 50 percent of the participants assigned as player C in a three-person dictator game, in which C decides between (575, 575, 575) and (900, 300, 600), sacrifice 25 tokens to equalize payoffs among players. This result favors the FS intuition of inequality aversion also in their choice situations.

unkind behavior. Therefore a natural prerequisite for games in the domain of reciprocity is that they are not merely one-sided decision tasks, as in the DG, but are of an interactive character.

Formal models of reciprocity have been developed by Rabin (1993), Falk and Fischbacher (2005), Dufwenberg and Kirchsteiger (2004), and CR. The latter two groups of authors combine reciprocity with inequality aversion and welfare concern respectively. Falk and Fischbacher (2005) show that there are parameter constellations for which their model is consistent with the stylized facts from the DG,[7] UG, PGG, MG, and some other games, but this result has to be contrasted with the complexity of their model. Without presenting any of formal model of reciprocity preferences in detail, the concept can be represented in an over-simplified form as:

$$u_i(x_i, x_j) = x_i + f_i(x_i, x_j) \cdot f_j(x_i, x_j) \ , \tag{2}$$

where f_i and f_j are so called kindness functions. f_i is a measure of how kindly player i treats player j, and f_j is a measure of player i's belief about whether player j is treating him kindly. A positive value for each of these kindness functions indicates kind behavior, and a negative value indicates mean behavior.

Any such utility specification assumes a tendency to reciprocate, since utility at a given payoff is maximized by rewarding an opponent who is believed to be kind with kind behavior. Likewise, if player i believes that player j was acting selfishly or mean he will wish to echo this behavior and react unkindly. In determining the reciprocal reaction it is crucial how a player interprets the behavior of other players, but the therefore necessary beliefs about other players' behavior are not captured by traditional game theory. For that reason reciprocity preferences require the framework of psychological game theory, as developed by Geanakoplos, Pearce, and Staccetti (1989). In this framework the equilibrium concept is, as in traditional game theory, a play of mutual best response, but requires in addition that the players' beliefs about the other players' actions match the equilibrium behavior. Due to this dependency, models of reciprocity have the drawback that their (psychological) equilibrium concept makes them often very difficult to apply, even in relatively simple experimental games. Sometimes there occur many equilibria due to different self-fulfilling beliefs about intentions, so that the theory then remains ambiguous in its prediction.[8]

[7] The ability of the Falk and Fischbacher (2005) model of reciprocity preferences to explain the DG is purely due to the inequality aversion part of their preferences.

[8] The survey "Theories of Fairness and Reciprocity - Evidence and Economic applications" by Fehr and

3.3 Altruistic Preferences and Welfare Concern

The Encyclopaedia Britannica (1998) defines an altruistic person as someone who feels the obligation "to further the pleasures and alleviate the pains of other people." "The hypothesis that people are altruistic has a long tradition in economics and has been used to explain charitable donations and the voluntary provision of public goods (see, e.g., G. Becker, 1974)." (Fehr and Schmidt 2003, p. 220) For an economic agent to be altruistic means that such a person's well-being is positively related to (increases with) other persons' well-being. Expressed formally, a person i is said to be altruistic if his utility function $u_i(x_1, ..., x_n)$ has the following properties:

$$\frac{\partial u_i(x_1,...,x_n)}{\partial x_j} > 0 \quad \text{if} \quad j \in [1, ..., i-1, i+1, ..., n] \; ; \tag{3}$$

i.e. the first partial derivatives of his utility function $u_i(x_1, ..., x_n)$ with respect to any co-players' payoff x_j, where $j \neq i$, are strictly positive. This easy representation makes it also straight forward to include an explicit altruistic inclination in any preference profile. To explicitly consider altruistic concerns one needs only to add the (weighted) payoffs of the other players to the utility function. Thus, a most simple model of a constant degree of altruism is given by:

$$u_i(x_1, ..., x_n) = x_i + \sum_{j \neq i} \gamma_j x_j \; . \tag{4}$$

If an altruist puts the same weight γ on the well-being of all others, then $x_i + \sum_{j \neq i} \gamma x_j \equiv (1-\gamma)\, x_i + \gamma \sum_{j=1}^{n} x_j$. This is equivalent to saying such a person is welfare concerned. Throughout the paper any agent i who cares to the same extent about others than for himself, i.e. all derivatives $\frac{\partial u_i(x_1,...,x_n)}{\partial x_{j \neq i}}$ equal $\frac{\partial u_i(x_1,...,x_n)}{\partial x_i}$, is called a perfect altruist. In model (4) perfect altruism implies $\gamma_i = 1$, for all $i \in \{1, ..., n\}$, and consequently, for perfect altruists only social welfare matters.

4 A Model of Inequality Aversion and Social Welfare (IASW)

As indicated by the generalized DG experiments welfare concern matters. Recognizing this it is an important step to combine welfare concern with inequality aversion. The combination

Schmidt (2003), and Gintis (2000)' textbook are comprehensive summaries, also of less well-known proposals in this field.

allows inequality aversion to be applied to a broader domain of games. Further, formalization of the hypothesis that these two motives interact yields rigorously testable model predictions, which can be confronted with experimental data and compared to competing approaches.[9] Doing so, I show in section 5.1 that the consistency findings of CR change in favor of an inequality-averse welfare maximizer.

While CR, who first accounted for welfare concern, combine reciprocity with a convex combination of the egalitarian and utilitarian welfare function:

$$W(x) = \delta \min \{x_1, ..., x_n\} + (1-\delta) \sum_k x_k \ , \tag{5}$$

part of this (broader) welfare-motive is readily represented in inequality aversion. Inequality aversion, when strong enough, exactly expresses the desire to equalize payoffs, i.e. it covers the egalitarian motive. Having recognized this, it is straightforward to include utilitarian welfare concern in an inequality aversion model. The utility function of player $i \in [1, ..., n]$ is then given by:

$$\begin{aligned} u_i(x) &= (1-\gamma_i)x_i + \gamma_i \sum_j x_j \\ &\quad -\alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_i - x_j, 0\} \ , \end{aligned} \tag{6}$$

where $\beta_i \leq \alpha_i$, $0 \leq \beta_i < 1$, and $0 \leq \gamma_i \leq 1$. Parameter γ_i expresses the degree of player i's welfare concern. The inequality averse welfare maximizer model retains the properties and interpretation of the FS inequality aversion model, but in addition player i directly cares for social welfare $\sum_{j=1}^n x_j$, whenever $\gamma_i \neq 0$. An increase in welfare, regardless of how it is distributed, now can have an offsetting effect on an increase in difference.

Rewriting (6) as follows gives the welfare maximization motive the altruistic interpretation:

$$u_i(x) = x_i + \gamma_i \sum_{j \neq i} x_j - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_i - x_j, 0\} \ . \tag{7}$$

Player i can thus be seen as a welfare concerned person (6), or equivalently an altruist (7). In both versions the model embodies a tension between inequality aversion and welfare concern, whenever these two goals have to be traded-off. The tension of the trade-off is

[9] Of course, arbitrary definition of preferences inevitably would lead to a theory that lacks any insight and is purely descriptive, but if new models are guided by experimental evidence and revised by experimental testing this caveat can be avoided.

determined by the weights $(\alpha_i, \beta_i, \gamma_i)$ of the different concerns, i.e. player-type dependent, and influenced by the specific exchange rate or multiplier vector $m = -\partial x_{-i}/\partial x_i$, where $x_{-i} = [x_1, ..., x_{i-1}, x_{i+1}, ..., x_n]$', at which welfare can be created.

To illustrate this, take a look at the utility function of a player who decides to give up one Euro in a two-player case. For the two-player case (7) simplifies to:

$$u_i(x) = x_i + \gamma_i x_j - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\} \ , \tag{8}$$

where $i \in \{1, 2\}$, $i \neq j$, and the marginal utility of player i is given by:

$$\frac{\partial u_i(x)}{\partial x_i} = \begin{cases} 1 + \alpha_i + [\alpha_i - \gamma_i]\, m & \text{if} \quad x_i < x_j \\ 1 - \beta_i - [\beta_i + \gamma_i]\, m & \text{if} \quad x_i > x_j \\ 1 - \gamma_i m & \text{if} \quad x_i = x_j \ . \end{cases} \tag{9}$$

Playing a standard DG with exchange ratio 1:1, i.e. $m = 1$, all player-types with $1 - 2\beta_i - \gamma_i > 0$ are utility-maximizing by sharing the pie equally. Since there are never players with $1 + 2\alpha_i - \gamma_i > 0$, even for welfare concerned players it is not optimal to volunteer more than half. Finally, only a perfect altruist, i.e. $\gamma_i = 1$, could be indifferent about giving a high amount.

It is worthwhile noting that, because of the linearity of the model, all crucial type characterizations (according to the algebraic sign of the respective marginal utilities) will yield linear parameter restrictions characterizing the player population. Continuing the example emphasizes this point. If we observe that 50 percent of the players offer zero and 50 percent offer 0.5 of the welfare, all we know is that -if our model is correct- there must be a player-type distribution function $F(\alpha, \beta, \gamma)$ such that 50 percent of the players lie below and 50 percent above the line $1 - 2\beta_i - \gamma_i = 0$ in the (β, γ)-plane of the parameter space. This implies that in the DG altruism γ and advantageous inequality aversion β are interchangeable within certain limits without affecting the observed outcome. Or, stating this result differently, a standard DG is suitable to characterize, but not to identify, the player-type distribution. Next, an extensive analysis of the IASW-model on the UG is presented.

4.1 Illustration of IASW on the Ultimatum Game

4.1.1 The Ultimatum Game

In the UG two players are allotted a sum of money (for simplicity normed to 1). The first player, often called the proposer, offers some share $\mathbf{s} \in [0, 1]$ of the money to the second player, called the responder. If the responder accepts, he receives what was offered, and the proposer retains the rest. If the responder rejects the offer, both players get nothing. Thus, for purely self-interest responders, accepting any offer is at least as good as rejecting and they will never reject. Self-interested proposers anticipate this behavior and then maximize their utility by offering nothing. Do not share and always accept is then the unique subgame-perfect equilibrium prediction for the UG.[10] The Ultimatum Game could not be simpler. It is thus an excellent starting point to test a new theory. Like the classical inequality aversion model, the IASW model assumes that offers $s > 0$ are triggered by inequality averse responders who would reject small offers. Independently of such triggered positive offers, non-zero offers will also be made by proposers who are themselves substantially inequality averse. The fact that the IASW responders may have an altruistic inclination mitigates their degree of inequality aversion, but does not change its essential implication that low offers could be turned down. For proposers with IASW preferences an initial inclination to share due to being inequality averse is reinforced by their degree of altruism.

figure 2 about here

4.1.2 The optimal Responder Behavior

Intuitively, where self-interest, altruism, and inequality aversion just offset each other, players are willing to share and accept a broad range of offers. The formal analyses of the UG starts with determining what share will by accepted by responders, assuming that they have IASW preferences. Proposition 1 summarizes the results.

[10] The responder's acceptance of a zero offer holds with weak preference. He is assumed to always accept. If the responder is assumed to always reject a zero offer the unique equilibrium becomes $(1 - x, x)$, where x is the smallest available money unit. Throughout, I use the first assumption.

Proposition 1 *It is a dominant strategy for the responder R to accept any offer*

$$s \equiv \begin{cases} s_H \in (0.5, 1] & \textit{always} & \textit{(i)} \\ s_M \in \{0.5\} & \textit{always} & \textit{(ii)} \\ s_L \in [0, 0.5) & \textit{if player-type is } 1 + 2\alpha_R - \gamma_R > 0 & \\ & \textit{and } s_L \geq \textit{max}\left\{\frac{\alpha_R - \gamma_R}{1+2\alpha_R - \gamma_R}, 0\right\} =: s_L^{Rc}(\alpha, \gamma) & \textit{(iii)} \\ s_L \in [0, 0.5) & \textit{if player-type is } 1 + 2\alpha_R - \gamma_R = 0 & \textit{(iv)} \end{cases}$$

and to reject s otherwise. [11], [12]

Proof. (i), (ii): A responder R will accept any offer $\tilde{s} \in \{s_M, s_H\}$ if his utility from accepting such an offer is strictly greater than his utility from a rejection, where social welfare is zero. Since $u_R(\tilde{s}) = (1 - \beta_R)\tilde{s} + (\beta_R + \gamma_R)(1 - \tilde{s}) > u_R(0) = 0$ for all responder-types an offer $\tilde{s} \geq 0.5$ will never be rejected. (iii), (iv): Similarly, acceptance of a low offer $s_L \in [0, 0.5)$ requires that s_L induces at least as much utility as a rejection. In terms of utility this means $u_R(s_L) = (1 + \alpha_R)s_L - (\alpha_R - \gamma_R)(1 - s_L) \geq u_R(0) = 0$. This inequality always holds when $1 + 2\alpha_R - \gamma_R = 0$, e.g. for all responders who are perfect altruists, without disadvantageous inequality aversion. Since the left hand side of the inequality is strictly increasing in small offers, for all other responder-types $1 + 2\alpha_R - \gamma_R > 0$ the inequality holds if low offers s_L are greater than a critical value $s_L^{Rc}(\alpha, \gamma)$. This critical value s_L^{Rc} is defined as the lowest offer that just fulfills the inequality, and is given by $s_L^{Rc}(\alpha, \gamma) = \frac{\alpha_R - \gamma_R}{1+2\alpha_R - \gamma_R}$ if $\alpha_R - \gamma_R \geq 0$ and by $s_L^{Rc} = 0$ if $\alpha_R - \gamma_R < 0$. Figure 4 shows the critical value $s_L^{Rc}(\alpha, \gamma)$ for different responder-types $(\alpha_R, \gamma_R) \in \{[0, 4] \times [0, 1]\}$. ■

figure 3 about here

The acceptance threshold $s_L^{Rc}(\alpha, \gamma)$ is strictly increasing in α_R, strictly decreasing in γ_R, and has an upper limit of $\lim_{\alpha \to \infty} s_L^{Rc}(\alpha, \gamma) = 0.5$ for all $\gamma_R \in [0, 1]$. Different responders

[11] Because of $\alpha_R > 0$ and $\gamma_R \in [0, 1]$, responder-types with $1 + 2\alpha_R - \gamma_R < 0$ do not exist.

[12] For notational simplicity subscripts of (α, β, γ) within the critical values for the responder s_L^{Rc} and the proposer s_L^{Pc} are dropped. The critical values can be uniquely identified by the superscripts R and P. In addition, critical values with (α, γ) dependence always represent the responder, while those with (β, γ) dependence refer exclusively to the proposer. This is due to the fact that, where the critcal values are relevant, i.e. for small offers $s_L < 0.5$, the analysis of proposer and responder behavior in the UG remains in specific inequality regions, namely the advantegeous and the inadvantegous respectively.

on one level curve share a common threshold $\hat{s}_L^{Rc}$ and their types are linearly related by $\gamma_R = \frac{(1-2\hat{s}_L^{Rc})\alpha_R - \hat{s}_L^{Rc}}{1-\hat{s}_L^{Rc}}$. All responders who are more altruistic and/or less inequality averse than the responders on the level curve given by $\hat{s}_L^{Rc}$ have lower acceptance thresholds and will also accept any offer $s = \hat{s}_L^{Rc}$.

4.1.3 The optimal Proposer Behavior

The optimal Proposer Behavior under Certainty about the Responder-Type

Proposition 2 *If a proposer P of type $(\alpha_P, \beta_P, \gamma_P)$ knew the responders preference-type $(\alpha_R, \beta_R, \gamma_R)$, then she will offer*

$$s^* \in \begin{cases} \left[s_L^{Rc}, 1\right] & \text{if } \alpha_P{=}\beta_P{=}0,\ \gamma_P{=}1 & (i) \\ \left[s_L^{Rc}, 0.5\right] & \text{if } 1-2\beta_P+\gamma_P=0 & (ii) \\ \{s_M\} = \{0.5\} & \text{if } 1-2\beta_P+\gamma_P<0 & (iii) \\ \left\{s_L^{Rc}\right\} & \text{if } 1-2\beta_P+\gamma_P>0 & (iv) \end{cases}$$

in equilibrium, and her offer s^* will always be accepted.

Proof. Acceptance of s^* follows directly from proposition 1. Claim (i), that proposers who are perfectly altruistic and not inequality averse will offer any share $s^* \in \left[s_L^{Rc}, 1\right]$, i.e. any share that is accepted, follows directly from their utility function, which allows them to always enjoy the social welfare $u_P(s^*) = (1-s^*) + s^* = 1$. For all other types, however, it is never optimal to offer a share $s_H > 0.5$. Or, put differently, there exists at least one offer $\tilde{s} \leq 0.5$ for which $u_P(\tilde{s}) \geq u_P(s_H)$. Rewriting $u_P(\tilde{s}) \geq u_P(s_H)$ and choosing $\tilde{s} = 0.5$ yields $u_P(s_H) - u_P(0.5) = (1+\alpha_P)(1-s_H) - (\alpha_P - \gamma_P)s_H - 0.5(1+\gamma_P)$, which can be simplified to $0.5(1+2\alpha_P-\gamma_P)(1-s_H)$, an expression that is < 0 for all proposer-types $(\alpha_P, \beta_P, \gamma_P)$, such that $[\alpha_P \neq 0,\ \beta_P \neq 0,\ \gamma_P \neq 1]$, whenever $s_H < 1$. For $s_H = 1$, $u_P(s_H) - u_P(0.5) < 0$ follows directly. Thus offering the share $\tilde{s} = s_M = 0.5$ is a dominant strategy for almost all proposers, and such proposers only need to compare their utility from offering a small share to their utility from offering a medium share when making a decision. Looking at this utility difference $u_P(s_L) - u_P(0.5) = (1-\beta_P)(1-s_L) + (\beta_P+\gamma_P)s_L - 0.5(1+\gamma_P) = 0.5(1-2\beta_P+\gamma_P)(1-s_L)$, regrouping yields the conditions $1-2\beta_P+\gamma_P \lesseqqgtr 0$ on (β_P, γ_P) for (ii), (iii) and (iv). Finally, noting that proposers who are not indifferent, but uniquely prefer a small offer will choose the small offer that maximizes their the utility, i.e. s_L^{Rc}, which is the lowest small offer that will still be accepted, completes the proof of proposition 2. ■

Under laboratory conditions the Ultimatum Game is played under complete anonymity. At the time making her offer the proposer does not know which type of responder she will face. She might, however, have a belief about the distribution of types which stems from past or outside experience she will apply when making her decision. How this uncertainty about the responder-type affects the optimal proposer behavior is discussed below.

The optimal Proposer Behavior under Uncertainty about the Responder-Type

Proposition 3 *If the proposer does not know the preferences of the responder, but believes that the responder-types* $(\alpha_R, \beta_R, \gamma_R)$ *are distributed according to a joint distribution function* $F(\alpha, \beta, \gamma)$*, then from her perspective the acceptance probability* $p(s)$ *of an offer* $s \in [0,1]$ *is given by:*

$$p(s) = \begin{cases} p_L = p(s_L) & (i) \\ p_M = p(s_M) = 1 & (ii) \\ p_H = p(s_H) = 1 & (iii) \end{cases},$$

$$\textit{where } p(s_L) = \begin{cases} 0 & \textit{if} \quad s_L < s_L^{Rc}(\underline{\alpha}, \overline{\gamma}) = \underline{s}_L^{Rc} \\ F(\alpha, \beta, \gamma \mid s_L \geq s_L^{Rc}) \in (0,1) & \textit{if} \quad \underline{s}_L^{Rc} \leq s_L < \overline{s}_L^{Rc} \\ 1 & \textit{if} \quad s_L \geq s_L^{Rc}(\overline{\alpha}, \underline{\gamma}) = \overline{s}_L^{Rc} \end{cases}$$

The critical values $\underline{s}_L^{Rc}$ *and* $\overline{s}_L^{Rc}$ *are the lowest and highest responder acceptance-thresholds for low offers* s_L *in the responder population.*

Proof. (i): Let $F(\alpha, \beta, \gamma)$ be the joint distribution function of responder-types $(\alpha_R, \beta_R, \gamma_R)$, with support $[\underline{\alpha}, \overline{\alpha}] \times [\underline{\beta}, \overline{\beta}] \times [\underline{\gamma}, \overline{\gamma}]$, where $0 \leq \underline{\alpha} < \overline{\alpha} < \infty$, $0 \leq \underline{\beta} < \overline{\beta} < \min\{1, \overline{\alpha}\}$ and $0 \leq \underline{\gamma} < \overline{\gamma} \leq 1$. p_L=0 and p_L=1 follow directly from proposition 1, part (iii) and (iv). $p_L = F(\alpha, \beta, \gamma \mid s_L \geq s_L^{Rc})$ is true by definition of $F(\alpha, \beta, \gamma)$ and $s_L^{Rc}(\alpha, \gamma)$. (ii) and (iii): p_M=p_H=1 follow directly from proposition 1, part (ii) and (i), respectively. ■

When considering her best offer in the game the utility maximizing proposer compares her expected utility $E[u_P(s)] = p(s) \cdot u_P(s)$ from different offers s, where the acceptance probability $p = p(s)$ is itself endogenous. Since p_M=p_H=1 the result (i) of proposition 2 still holds: Only perfectly altruistic and non inequality-averse players are weakly willing to propose a high offer. Therefore for all other proposer-types $\{(\alpha_P, \beta_P, \gamma_P) \mid (\alpha_P, \beta_P, \gamma_P) \notin (0,0,1)\}$, who already strictly prefer a medium to a large offer, it is sufficient to compare the expected utility maximum from the best low-offer with their utility from a medium offer. Given the responder-type distribution $F(\alpha,\beta,\gamma)$, thus $p(s)$, let $s_L^* \in [0,0.5)$ be the best low-offer a proposer can make; i.e. $s_L^* \in \arg\max_{s_L} \{E[u_P(s_L)] = p(s_L)\cdot u_P(s_L)\}$. To find a formal choice-condition for the proposer, I compare the expected utility $E[u_P(s_L^*)]$ from making the best low-offer to the utility this proposer-type will gain from a medium offer $u_P(s_M)$.

This comparison leads to an *offer-choice-condition*. The offer-choice-condition is a set of critical proposer values $s_L^{Pc}(\beta,\gamma|p_L^*)$, that depend on the proposer-type (β_P,γ_P) as well as on the predetermined[13] best-low-offer acceptance probability $p_L^* = p(s_L^*)$:

$$\begin{aligned} E[u_P(s_L^*)] &= p_L^* \cdot u_P(s_L^*) \leq u_P(s_M) \Leftrightarrow \\ &\Leftrightarrow p_L^* \cdot \{(1-s_L^*) + \gamma_P s_L^* - \beta_P[(1-s_L^*) - s_L^*]\} \\ &\leq (1-s_M) - \gamma_P s_M = 0.5(1+\gamma_P) \\ &\Leftrightarrow p_L^* \cdot [(1-\beta_P) - s_L^*(1-2\beta_P-\gamma_P)] \leq 0.5(1+\gamma_P) \end{aligned}$$

$$\Leftrightarrow \begin{cases} s_L^* \geq \frac{1-\beta_P-0.5(1+\gamma_P)(p_L^*)^{-1}}{1-2\beta_P-\gamma_P} =: s_L^{Pc}(\beta,\gamma) & \text{if} \quad 1-2\beta_P-\gamma_P > 0 \qquad \text{(A)} \\ s_L^* \leq s_L^{Pc}(\beta,\gamma) & \text{if} \quad 1-2\beta_P-\gamma_P < 0 \qquad \text{(B)} \\ \text{always,}^{14} & \text{if} \quad 1-2\beta_P-\gamma_P = 0 \\ \qquad \text{where } E[u_P(s_L^*)] < u_P(s_M) \text{ for all } p_L^* \in [0,1) & \text{(C1)} \\ \qquad \text{where } E[u_P(s_L^*)] = u_P(s_M) \text{ for } p_L^* = 1\ . & \text{(C2)} \end{cases} \tag{10}$$

All proposers choose their preferred offer from $\{s_L^*, s_M\}$ according to the respective part of this offer-choice-condition. For all non-perfectly altruistic or inequality averse proposer-types, this preferred offer will also be their optimal offer s^*. To illustrate the offer-choice-condition figure 4 presents a simulation of the critical proposer-value $s_L^{Pc}(\beta,\gamma|p_L^*=0.7)$ and the proposer-types' choices over the admitted parameter range $\{(\beta,\gamma)\in[0,1)\times[0,1]\}$ for an -arbitrarily assumed- optimal acceptance-probability p_L^*=0.7. Where the critical proposer value $s_L^{Pc} \notin [0,1]$ takes the value of an undefined offer, its value is plotted as the nearest limit of the integral $[0,1]$, i.e. the closest defined offer. All critical thresholds in the proposer-type characterization are lines in the (β,γ)-plane, where $s_L^{Pc}=0$. The plane $s_L^{Pc}=0.5$ in figure

[13] p_L^* is predetermined as proposers at this stage already have chosen their best low-offer s_L^* from all available low offers.

4 shows their projection. The first characteristic line $1-\beta_P-0.5(1+\gamma_P)\left(p_L^*\right)^{-1}$ [1] exists for all $p_L^*>0.5$, and converges with increasing p_L^* to the second line $1-2\beta_P-\gamma_P=0$ [2], which occurs at the discontinuity of the proposers' critical value $s_L^{Pc}(\beta,\gamma)$. Note that $1-2\beta_P-\gamma_P>0$ to the left of line [2] and <0 to its right. Thus, by (10A), if the best low-offer s_L^* in the region $1-2\beta_P-\gamma_P>0$ to left of [2] is weakly bigger than s_L^{Pc}, proposers will make a low offer. If not, i.e. $s_L^*\in\left[0,s_L^{Pc}\right)$, they choose the medium offer s_M. In the region to the right of [2] part (10B) of the offer-choice-condition is valid and the opposite is true.

The non-existence of the first threshold-line [1] for $p_L^*\leq 0.5$ is intuitive. If the acceptance probability of low offers is small, i.e. $p_L\leq 0.5$ for any s_L, then already the expected utility of the lowest low-offer, $s_L=0$, is never higher than the utility from offering half of the welfare, which will be accepted with certainty. On the contrary, for higher acceptance probabilities of low offers $p_L>0.5$, as they are observed in experiments, it becomes worthwhile for some proposer-types to offer small shares instead of an equal division.

figure 4 about here

With the offer-choice-condition I can fully characterize the proposer-type distribution. If $p_L^*>0.5$, by (10B), proposers who possess relatively low (β,γ)-values, i.e. the more self-interested types, prefer to make any low-offer up to their critical value $s_L\in\left[0,s_L^{Pc}\right)$. These proposers will choose the lowest $s_L=s_L^*$ that suffices to yield the predetermined acceptance probability p_L^*. If such a low offer is not available, i.e. $s_L^*\geq s_L^{Pc}$ they will choose s_M. As inequality aversion and/or altruistic concern grow there is a turning point, that is $1-2p_L^*(1-\beta_P)+\gamma_P>0$ (all proposers to the right of [1]), after which all proposer-types strictly prefer to make a medium offer. Their degree of altruism combined with their relatively high inequality aversion is now so strong that they do not want to propose any risky, unequal offer, even though they could ensure themselves a higher payoff. However, there are also proposers (to the right of [2]) that are so highly inequality-averse and altruistic that regardless of the responder-type distribution, and hence the offer acceptance-probability, they always want to offer s_M. Finally, there are potentially non-altruistic and inequality-averse proposers that are indifferent between a range of low offers that will be accepted for sure and a medium offer, i.e. they choose $s^*\in\left[\overline{s}_L^{Rc},s_M\right]$. (10C2) They exist only, when their favored set of low offers

$\check{S}_L^*$ contains exclusively low offers $\check{s}_L^* \in \check{S}_L^*$, which induce a low-offer acceptance probability $p(\check{s}_L^*) = 1$; i.e. the lowest offer in this set must already be accepted with certainty. These proposers-types can be found on the line [2], that for them coincides with [1]. For those proposers the utility gained by giving away one Euro just offsets the marginal disutility from loosing this Euro because of their altruistic capability of enjoying others' well-being and the reduced difference. Proposition 4 collects these findings about the optimal offer s^* under uncertainty about the responder-type when only the type-distribution is known.

Proposition 4 *The optimal offer s^* of a proposer depends on her own type $(\alpha_P, \beta_P, \gamma_P)$ and the responder-type distribution $F(\alpha, \beta, \gamma)$. The responder-type distribution $F(\alpha, \beta, \gamma)$ determines the offer acceptance-probability $p(s)$. According to whether a proposer-type facing $F(\alpha, \beta, \gamma)$ has already a best low-offer s_L^* that is big enough to induce a low-offer acceptance probability of $p(s_L^*) = 1$, two cases for the optimal offer s^* can be distinguished:*

(A) If $s_L^* \geq \overline{s}_L^{Rc}$, then $p(s_L^*) = 1$ and the optimal offer is given by:

$$s^* \in \begin{cases} \left[\overline{s}_L^{Rc}, 1\right] & \text{if} \quad \alpha_P = \beta_P = 0,\ \gamma_P = 1 & (i) \\ \left[\overline{s}_L^{Rc}, s_M\right] & \text{if} \quad 1 - 2\beta_P - \gamma_P = 0 & (ii) \\ \{s_M\} & \text{if} \quad 1 - 2\beta_P - \gamma_P < 0 & (iii) \\ \{s_L^*\} & \text{if} \quad 1 - 2\beta_P - \gamma_P > 0 & (iv) \end{cases}$$

(B) If $s_L^* < \overline{s}_L^{Rc}$, then $p(s_L^*) < 1$ and the optimal offer is given by: [15]

$$s^* \in \begin{cases} [s_M, 1] & \text{if} \quad \alpha_P = \beta_P = 0,\ \gamma_P = 1 & (v) \\ \{s_M\} & \text{if} \quad 1 - 2p_L(1-\beta_P) + \gamma_P \geq 0 & (vi) \\ \{s_M\} & \text{if} \quad 1 - 2p_L(1-\beta_P) + \gamma_P < 0 \text{ and } s_L^* > s_L^{Pc} & (vii) \\ \{s_L^*\} & \text{if} \quad 1 - 2p_L(1-\beta_P) + \gamma_P < 0 \text{ and } s_L^* \leq s_L^{Pc} & (viii) \end{cases}$$

Proof. (A): Since these proposer-types have a most preferred low offer s_L^*, which is higher than the highest low-offer acceptance threshold in the responder population $\overline{s}_L^{Rc}$, all responders would accept the low offer s_L^*, if such a low offer was made. Thus, the proposer's offer-choice problem under uncertainty about the responder-type in case (A) eventually follows her offer-choice problem under certainty about the responder-type, in which only offers that will be surely accepted are made. Hence the proof of (i)-(iv) is analogue to the proof of proposition 2(i)-(iv).

[15] Subcases (vii) and (viii) are only feasible for $p_L \geq 0.5$.

(B): Keeping figure 9a in mind illustrates the single steps of the proof of proposition 4B. (v) If $\tilde{s}^* \in [s_M, 1]$, by proposition 3 (ii) and (iii), $p(\tilde{s}^*) = 1$, and thus the proof of this part is analogue to the proof of proposition (2ii). (vi-a) Subregion, where $1 - 2\beta_P - \gamma_P < 0$. Firstly, since $p_L^* \in [0, 1]$,

$$\begin{aligned}\lim_{(\beta,\gamma)\to(1,1)} s_L^{Pc}(\beta,\gamma) &= \lim_{(\beta,\gamma)\to(1,1)} \frac{1-\beta_P - 0.5(1+\gamma_P)\left(p_L^*\right)^{-1}}{(1-2\beta_P-\gamma_P)} = \\ &= \frac{1-1-0.5(1+1)\left(p_L^*\right)^{-1}}{(1-2-1)} = \frac{1}{2p_L^*} \geq 0.5.\end{aligned}$$

Secondly, since $1 - \beta_P - \frac{1+\gamma_P}{2p_L^*} < 0 \; \forall \; 1 - 2\beta_P - \gamma_P < 0$ and $p_L^* \in [0, 1]$,

$$\lim_{(\beta,\gamma)\to(1-2\beta_P-\gamma_P=-0)} s_L^{Pc}(\beta,\gamma) = \frac{1-\beta_P - 0.5(1+\gamma_P)\left(p_L^*\right)^{-1}}{-0} = \infty.$$

Thirdly, s_L^{Pc} is a continuous function of β and γ in the regions $1 - 2\beta_P - \gamma_P \lesseqgtr 0$. Hence, in the region $1 - 2\beta_P - \gamma_P < 0$ the critical value for the proposer s_L^{Pc} is always (weakly) bigger than 0.5 for all allowed (β, γ)-combinations and any low-offer acceptance probability. Since $s_L < 0.5$ by (10B) it thus follows that these proposers strictly prefer a medium offer to any small one. (vi-b) Subregion 2, where $1 - 2\beta_P - \gamma_P = 0$. By (10C1) s_M is strictly preferred. (vi-c) Subregion 3, where $1 - 2\beta_P - \gamma_P > 0$ and $1 - 2p_L^*(1 - \beta_P) + \gamma_P \geq 0$. If $1 - 2\beta_P - \gamma_P > 0$ and $1 - 2p_L^*(1 - \beta_P) + \gamma_P \geq 0$ do not coincide, the critical proposer value s_L^{Pc} in the interjacent region is negative. This can be formally proved, or seen from simulations (as exemplified for p_L^*=0.7) and the continuity properties of s_L^{Pc}. Thus any low offer is greater than the negative s_L^{Pc} and from (10A) it follows directly that the medium offer s_M is preferred; with strict preference for all proposer-types $1 - 2p_L^*(1 - \beta_P) + \gamma_P > 0$. (vii) and (viii) In the region $1 - 2p_L(1 - \beta_P) + \gamma_P < 0$ the critical proposer value s_L^{Pc} is positive and further $s_L^{Pc} \leq 0.5$. Thus by (10A) the best small offer s_L^*, that just induces the previously determined best low-offer acceptance probability $p_L(s_L^*)$, is strictly preferred by these proposers to s_M, whenever $s_L^* < s_L^{Pc}$ and vice versa. (Indifferent types are assumed to choose s_L^*.) ■

The IASW implications, which are collected in propositions 1 and 4, account for many of the facts observed in the Ultimatum Game. They indicate that large offers above 0.5 are always accepted, but are only potentially made by non-inequality averse perfect altruists. Very low offers are likely to be rejected by inequality averse types, and thus a variety of proposers considers low and medium offers. This behavior is encouraged as inequality aversion or welfare concern for them increases. As the FS model, IASW preferences have the potential to fit the UG well. By calibrating the IASW model on stylized facts from the UG, I next characterize the distribution of the heterogeneous player-types.

4.2 Calibration of the IASW Model

Fehr and Schmidt (1999) stylized the actual offers observed in several Ultimatum Games as follows: "There are roughly 40 percent of the subjects who suggest an equal split. Another 30 percent offer $s \in [0.4, 0.5)$, while 30 percent offer less than 0.4. There are hardly any offers below 0.25." They further note that in all UG experiments there is a fraction of subjects who reject offers even if they are very close to an equal split and thus "conservatively" assume that 10 percent of the subjects have an acceptance threshold of $s_L^{Rc} = \frac{4}{9}$. They consider that another, typically much larger, fraction of the population (about 30 percent) insists on getting at least one-third of the welfare, and that another 30 percent of subjects insist on getting at least one-quarter. Finally, as they observe, the remaining 30 percent of subjects are happy to accept any positive offer.

If a proposer does not know the type of her opponent, but believes that the acceptance probability of an offer $p(s)$ is given by this distribution of the acceptance thresholds $s_L^{Rc}(\alpha, \gamma)$, it is straightforward to compute her optimal offer $s^* \in \arg\max_s \{E[u_P(s)] = p(s) \cdot u_P(s)\}$ as a function of her type. (see appendix AI, p. 33) The optimal offer is then given by:

$$s^* \in \begin{cases} \left[\frac{4}{9}, 1\right] & \text{if} \quad \alpha_P{=}\beta_P{=}0,\ \gamma_P{=}1 \\ \left[\frac{4}{9}, s_M\right] & \text{if} \quad 1 - 2\beta_P - \gamma_P = 0 \quad \text{and} \quad (\alpha_P, \beta_P, \gamma_P) \notin \{(0,0,1)\} \\ \{s_M\} & \text{if} \quad 1 - 2\beta_P - \gamma_P < 0 \\ \{s_L\} = \left\{\frac{4}{9}\right\} & \text{if} \quad 1 - 2\beta_P - \gamma_P > 0 \quad \text{and} \quad 1 - \frac{17}{4}\beta_P - \frac{13}{4}\gamma_P > 0 \\ \{s_L\} = \left\{\frac{1}{3}\right\} & \text{if} \quad 1 - 2\beta_P - \gamma_P > 0 \quad \text{and} \quad 1 - \frac{17}{4}\beta_P - \frac{13}{4}\gamma_P \leq 0 \end{cases}.$$

All the above information, and the thereby implied characterization of the player-type distributions that are suitable to reproduce the stylized equilibrium offers s^*, are summarized in the following table:

Table 1: *Characterization of Parameter Distributions, Responders' Critical Values, and Equilibrium Offers according to FS' Stylized Facts*

observed	**calculated**
distribution of acceptance thresholds and corresponding responder-types[16]	cumulative acceptance probability
30% $s_L^{Rc} = 0 \Leftrightarrow \alpha \leq \gamma$	30% $p_L(0) = 0.3$
30% $s_L^{Rc} = \frac{1}{4} \Leftrightarrow 1 - 2\alpha + 3\gamma = 0$	60% $p_L(\frac{1}{4}) = 0.6$
30% $s_L^{Rc} = \frac{1}{3} \Leftrightarrow 1 - \alpha + 2\gamma = 0$	90% $p_L(\frac{1}{3}) = 0.9$
10% $s_L^{Rc} = \frac{4}{9} \Leftrightarrow 1 - \frac{1}{4}\alpha + \frac{5}{4}\gamma = 0$	100% $p_L(\frac{4}{9}) = 1$

observed	**assumed**
equilibrium offer (optimal offer)	suggested distribution of proposer-types[17]
30% $s^* = \frac{1}{3}$	30% $1 - \frac{17}{4}\beta_P - \frac{13}{4}\gamma_P \leq 0$
30% $s^* = \frac{4}{9}$	60% $1 - 2\beta_P - \gamma_P > 0$
40% $s^* = 0.5$	40% $1 - 2\beta_P - \gamma_P < 0$
0% $s^* = s_H$	0% $\{\alpha_P = \beta_P = 0, \gamma_P = 1\}$

Figures 5a and 5b below show how IASW allows welfare concern to interact with inequality aversion. The influence of welfare concern γ_i, $i \in \{R, P\}$, is graphed for responders R and proposers P against the respectively relevant weights of inequality aversion α_R and β_P. A separation of inequality regions according to proposers and responders is possible since in the stylized facts no offers greater than half of the endowment occur.

figure 5a and 5b about here

The figures also illustrate as to how the FS-distribution of thresholds changes when altruism comes into play, i.e. $\gamma_i > 0$. If one would like to accommodate a share of 30 percent

[16] Restrictions follow directly from Proposition 1.

[17] For a formal derivation of these restrictions see appendix AI. (p. 33)

of perfect altruists with $\gamma_i = 1$ in the population then the FS-distribution of thresholds changes, for instance, to a IASW parameter distribution as presented in table 2. The UG play prediction under both preferences, the FS-model and the richer IASW-model, remains equivalent.

Table 2: *Assumptions about the Distribution of Preferences*

FS (p. 844)				IASW example					
α		β		α	γ		β	γ	
0	30%	0	30%	0.5	0	30%	0	0	30%
0.5	30%	0.25	30%	1	0	30%	0.25	0	30%
1	30%	0.6	40%	1	1	**30%**	0.6	0	10%
4	10%			4	0	10%	0.6	1	**30%**

The depicted IASW parameter distribution is a sample from the continuum of distributions that allow for a positive degree of altruism in the UG. Nonetheless, focussing on perfect altruists is an intuitive starting point to get a grasp on the more diverse variety of player-types under the IASW model.

Let's have a closer look at the interaction between inequality aversion and altruism that is implied by the IASW model. In figure 5a the tension between the two motives in the responder case implies that responders with a common acceptance threshold sit on a common (upwards sloping) line, and if some of them are more altruistic than others this has to be offset by a higher aversion to being on the short end of the payoff distribution. Without this offsetting effect responders would be willing to accept lower offers like their less altruistic colleagues as their altruistic inclination increases.[18] Conversely, for proposers (figure 5b) inequality aversion and welfare concern reinforce each other as motives to share their endowment. Since, ceteris paribus, more concern for others implies a higher willingness to share, proposers who make equal offers can (and need) to be less inequality averse. Thus, the critical type restrictions are all downwards sloping. Because only a discrete distribution of acceptance thresholds exists, proposers offering the same share are in regions that collect all proposer-types with an equal or higher willingness to share until the next acceptance threshold is available. Disregarding the relevance of welfare concern, i.e. $\gamma = 0$ for all

[18] Since there is no negative offer available, it is intuitive why the restriction for 30 percent accepting zero offers does not need to be binding; i.e. these responder-types can also be to the left of the line $\alpha_R = \gamma_R$.

players, the IASW model coincides immediately with the traditional FS model and the (α, β)-parameter distribution found by FS is represented on the abscissae.

The coherence of the models makes the UG a very good starting point to introduce the IASW model, because the UG was extensively analyzed by Fehr and Schmidt (1999). However, in the UG welfare is only a binomial matter of existence and thus the game does not lend itself to much quantitative exploration of welfare concern. Any (α, β, γ)-distribution fulfilling the above restrictions on proposer and responder-types is suitable to reproduce the observed stylized facts of the UG in the IASW model, and there is no need to involve welfare concern; i.e. (α, β, γ)-distributions with $\gamma_i = 0 \; \forall \; i \in [1, ..., n]$ are feasible. This will change in the next section, where the IASW model is confronted with welfare relevant experimental data. Nonetheless, making the assumption that all proposers and responders in the UG are randomly drawn from one representative population, it is already possible to infer one upper-bound γ-restriction from the given data.[19] Namely, at least 30 percent of the proposers are not more than moderately welfare concerned, i.e. $\gamma \leq \frac{4}{13}$ (see figure 5b). Remarkably, this result is roughly consistent with experiments from Andreoni and Miller (2002), who found that welfare concern matters and suggested that 47.2% (22.7%) of the participants behave as if they were purely self-interested, 30.4% (14.2%) according to Leontief and 22.4% (6.2%) according to Utilitarian preferences [weak types (strong types)].[20]

5 The IASW Model on Welfare Relevant Data

5.1 Preference Evaluation by Charness and Rabin (2002)

A primary intention of CR is to collect evidence indicating that the "apparent adequacy of inequality aversion models has likely been an artefact of powerful and decisive confounds in the games used to construct these models." (CR, p. 849) In order to avoid framing influence of the game structure, CR construct an abstract experiment, in which subjects were asked to make choices over different payoff bundles. (see figure 6) These choice situations were either framed in a dictatorial manner, in which player B had to choose between $(x_{A,}^{left}, x_{B}^{left})$ vs. $(x_{A,}^{right}, x_{B}^{right})$, or as a sequential choice, in which first a player A chooses between

[19] Further assuming that the population of players is sufficiently large, such that the acceptance probability does not change by randomly drawing one player from the population for the proposer role.

[20] These preference-types translate to game specific (α, β, γ)-restrictions of the IASW model, such that IASW actors behave in a game as if they had one of these classical types of preferences.

$(x_{A,}^{end}, x_B^{end})$, in which case the game ends, or letting B enter, in which case B could choose between two given allocations $(x_{A,}^{left}, x_B^{left})$ vs. $(x_{A,}^{right}, x_B^{right})$.[21]

figure 6 about here

The experiment was conducted in thirteen sessions with a total of 32 games and 1680 observations. Eight sessions (games 1 to 12) were run at the Universitat Pompeu Fabra in Barcelona and 5 sessions (games 13 to 32) at the University of California at Berkeley. No participant could attend more than one session. According to CR (p. 825) "average earnings were around $9 in Barcelona and $16 in Berkeley [or] about $6 and $11, net of the show-up fee paid." In the sessions for games 17-32, players were anonymously rematched with different partners for up to six games, where games also included hypothetical choices. In the other sessions, comprising games 1-16, candidates were only invited to play once. Following CR's consistency analysis of preferences, I disregard five three-player games comprising 106 observations. The payoff allocations offered and actual choices made in the 27 two-player games which I consider are reported in the appendix AII (p. 34).

Based on missing evidence for positive, but significant evidence for negative, reciprocity in the data: *A choice by A that favored B's payoff situation, i.e.* $x_B^{end} < \max\left\{x_B^{left}, x_B^{right}\right\}$ *was not rewarded, but an unkind action by which A entered the game when* $x_B^{end} > \max\left\{x_B^{left}, x_B^{right}\right\}$, *triggered punishment*,[22] CR propose a model of egalitarian-utilitarian welfare concern and a negative reciprocity motive. As indicated by a consistency check, these so-called social-welfare preferences have the potential to fit the data much better than FS' model of inequality aversion and a self-interest model. Consistency thereby requires that the observed experimental behavior is explainable with a preference type for any parameter-value vector permitted by the parameter-space restrictions of each type of (social) preference. Table 3 below shows the consistency numbers reported by CR. Clearly, social-welfare preferences with an overall consistency of 95 percent outperform inequality aversion that achieves a consistency of 86 percent, and the neoclassical model of pure self-interest with a consistency

[21] CR (p. 830 et sqq.) also remark that "it is of course somewhat arbitrary to compare models on this set of games" and point out that it is difficult to "define a 'fair' test of the different [...] preferences because we do not know the most appropriate array of games to study."

[22] CR find these results by comparing responder games to their dictatorial counterparts.

of 84 percent. Since CR argue that to evaluate the consistency of A-players' choices it is unclear whether one should assume that A estimates the probability of B's subsequent decision correctly, they report consistency for both cases.[23]

Table 3: *Consistency Results over the 27 Choice Games According to CR*

	total # of observations	self-interest	social-welfare	inequality aversion	perfect altruism[24]	IASW
1	232	68%	97%	75%	66%	100%
2	903	79%	91%	76%	67%	97%[25]
3	671 (with)	69%	97%	90%	53%	98%[26]
	671 (w/o)	94%	99%	100%	65%	100%
4	**1574 (with)**	**73%**	**94%**	**82%**	**61%**	**98%**
	1574 (w/o)	84%	95%	86%	66%	98%

Note: (1) B's behavior in the seven dictator games, (2) B's behavior in all games, (3) A's behavior in the 2-person games with(out) assumption: A's belief about B is correct, **(4) Consistency of all choices over all games with**(out) **assumption on A´s belief**

While I like CR's novel idea of challenging competing preferences on very simple decision situations, in my perception their set of decision games lay outside the domain of a model of pure inequality aversion. Welfare, which can be an independent behavioral motive, was usually directly concerned, and thus their games constitute an environment in which IASW preferences should be tested. This appears even more compelling when recognizing that in the original preference-test a direct relevance of welfare is also considered by CR, who

[23] In the light of the search for a common parameter-distribution that belongs to a correct model, one unique distribution should be able to explain several observation across games. Thus going one step further, assuming that such a distribution is common knowledge, i.e. A estimates the probability of an subsequent action correctly, is a sensible assumption.

[24] Consistency of the choices is tested with the model of altruism (4) for the two player case and $\gamma_i = 1$ for $i \in \{1, 2\}$.

[25] Consistency violations occur in games 11, 22, 28, 32 by a total number of $4 + 1 + 11 + 9 = 25$ B-players. Allowing for indifference in games 28, 32 to explain the observed behavior, consistency increases to 99% for B's behavior and to 100% (99%) for the overall consistency respectively.

[26] Consistency violations occur in games 1, 13, 31 by a total number of $2 + 3 + 7 = 12$ A-players.

favor a hybrid model. Their social-welfare preferences include the welfare motive (5) as well as reciprocity. If extending the consistency check, which was chosen as a mean of evaluation, to the IASW model, then IASW preferences spontaneously outperform the other types of (social) preferences, including CR's social-welfare preferences. The proportion of observations compatible with the IASW-model is higher than the proportion compatible with all other types of preferences. IASW suggest an almost perfect fit of the data with an overall consistency measure of 98 percent.[27] As a further benchmark I also analyzed and report consistency for a model of perfect altruism. In the seven DGs the explanatory potential of perfect altruism (66 percent) comes close to that of narrow self-interest (68 percent). Otherwise, however, perfect altruism is clearly outperformed.

5.2 Excursion: Flaws in the Consistency Measure

There is an important point to note about the comparison undertaken by CR. The reported consistency numbers for all types of preferences are based on the idea that a preference profile is able to fit the data with any set of permitted parameter values. Thus the consistency values do not express the real fit of the respective preference models but only a potential fit. Consequently, they can only be seen as an upper explanatory success bound. To achieve the best fit of the data, it is necessary to find one coherent parameter distribution over the whole player population which is consistent with each single restriction implied by any game played within the population. A short example illustrates this point and shows that within the CR data for IASW preferences this is not possible. Consider games 14, 18, and 26 from the CR data:

[27] In comparision to pure inequality aversion, IASW buys its consistency gain through the additional parameter for welfare concern. As can be seen in Table 3 (p. 24), consistency increases between 8-25 percent are directly attributable to γ_i-values >0. IASW with three parameters, nonetheless, is specified more parsimoniously than CR's social-welfare preference with five parameters.

Table 4: *CR - Some Game by Game Results*

game	p^{end}	p^{enter}	p^{left}	p^{right}
two-person dictator games:				
26: B chooses $(0,800)$ vs. $(400,400)$			.78	.22
two-person response games:				
14: A chooses $(800,0)$ or lets				
B choose $(0,800)$ vs. $(400,400)$	.68	.32	.55	.45
18: A chooses $(0,800)$ or lets				
B choose $(0,800)$ vs. $(400,400)$	.00	1.00	.56	.44

Assuming IASW preferences the decision problem "B chooses $(0,800)$ vs. $(400,400)$" implies, irrespective of its A-move history, the same restriction on B-player-types in the (β,γ)-parameter space: $u_B(0,800) \lesseqqgtr u_B(400,400) \Leftrightarrow 1+2\beta-\gamma \lesseqqgtr 0$ [henceforth +]. The (β,γ)-combinations on the critical threshold-line, where [+] holds as equality, correspond to the marginal B-players who are indifferent between choosing left or right. Player-types in the lower half of the (β,γ)-space $[0,1)\times[0,1]$, i.e. underneath the critical line, correspond to the Bs choosing right. B-types in the upper part choose left. A graphical representation of these two distinguishable B-type regions is plotted below:

figure 7 about here

Since the history of the game does not influence agents with IASW preferences, the decision problems 14, 18, and 26 are equivalent from B-players' perspective, i.e. B-decisions are independent of the preceding A-decision. Therefore without any hint for a systematic difference between the player groups invited for these games,[28] IASW implies similar proportions of Bs choosing left and right in all three games. This implication is refuted by the CR data: Calibrating the IASW model on the choices observed in game 26 requires that 78 percent of the B-players have (β,γ)-values above and 22 percent (β,γ)-values below the critical line [+]. Games 14 and 18, on the contrary, both suggest that only about 55 percent of the B-types lie above [+] and 45 percent below [+]. This difference in the observed proportions of

[28] CR's experimentees were randomly invited students. Cultural differences are assumed not to matter.

the left/right-choices between the two-person dictator game 26 and the two-person response games 14 and 18 is statistically significant at the 1 percent level.[29]

Thus, clearly these three games are not reconcilable with one single parameter distribution, even though the IASW model is fully consistent with each of the three single games.[30] In other words the extremely high consistency value of the IASW model for the whole data set cannot be achieved with a unique player-type distribution and the true fit for the data lies below that value. This aspect is the more severe since CR report further history-dependent inconsistencies in likewise equivalent subgames that point against a unique B-player distribution. Using a proportion test they find that on average B-players' decisions differ significantly between games (7) and (29), (3, 4, 21) and (2, 17), (3, 4, 21) and (10, 2, 17), (2) and (1), as well as (17) and (13).[31] However, they also remark that due to their particular experimental design, in which players partly make hypothetical decisions, it is likely that they overestimate the significance of these differences, since their observations are not independent. Nonetheless, when accounting for the possible dependence by reducing the dataset to the number of independent and actual choices, CR report that an indication for a systematic influence of the game history on B's decision making behavior is still supported by the data, at least for some games. Thus the criticism against the consistency measure is robust.

Exploring to what extent CR's consistency measure overestimates the true fit for each preference would require calibration of each model. For the IASW model the important steps of the calibration procedure and some implications of a calibrated model are sketched below. There was no other attempt to calibrate any other social preference on this data, especially not by CR for their social-welfare preferences. It thus remains unresolved to what extent the evaluation of each preference is affected. Since preferences of pure inequality aversion are a special case of IASW, inequality aversion faces the problem as described

[29] CR use a difference in proportions test (see Charness and Grosskopf 2004) and report p-values of 0.01 for the null-hypotheses B's allocative decision is equal across games 14 and 26, as well as 18 and 26 respectively.

[30] CR find their observations in these three games somewhat puzzling. They do not support any concept of reciprocity, since the change in B's behavior goes in the same direction, regardless of whether the history is of kind or unkind behavior. In 14, any A who enters can be understood as a player who commits a friendly action, which is rewarded by B through a significantly higher number of equal splits (left-choices). However, in game 18, where A commits from B's perspective a rather unkind action by entering, subsequent B behavior changes in the same way. Consequently, it might not be reciprocity, but rather a fixed effect that is at work. CR report other game pairs, in which similar interpretation of A's move in terms of kindness induced significant changes of B's behavior, that can be clearly interpreted as negative reciprocity.

[31] At the 1%, 15%, 10%, 1% and 1% significance level, respectively.

above. Social-welfare preferences, however, are partly motivated by the history specific systematic differences of, on the B-player level, equal decisions. Their consistency measure might therefore not overestimate the true fit for the same reasons, but may be flawed in another, here unexplored way. Eventually, the consistency measure and this specific problem should not be overvalued in a discussion about preferences, since its adequacy also depends upon whether a model's parameter restrictions have an intuitive interpretation, whether implied parameter-value constellations are realistic, and upon how well-known systematic behavioral patterns are *ex ante* intentionally covered by a social preference.

This notwithstanding, I believe that three important aspects can be derived from the consistency results. First, in contradiction of the findings of CR inequality aversion is a powerful concept, and has the potential to explain their data (to a much greater extent than it seemed at a first glance) once it is extended by welfare concern. Second, with their own measure of comparison -which is however questionable- the, in this context adequate, IASW-model is a model of inequality aversion that outperforms the reciprocity and welfare combination of social-welfare preferences. Third, parts of CR's analysis are based on a consistency check that can only be understood as an upper confidence bound, so that any conclusion drawn on the consistency measure is problematic.

5.3 Calibration of IASW on the Charness and Rabin (2002) Data

Each choice situation of the CR data implies a linear restriction in the player-type space. Exemplifying calculations are provided in appendix AIII. (p. 35) All restrictions are planes in the (α, β, γ)-space which separate the type-space in two regions, that collect all player-types that potentially would have chosen the same way, if asked to enter the particular game. Because observations of individual participants are from at most seven games, preferences on the individual level can only be vaguely characterized, by collecting the type-restrictions for the single players over games. However, the aggregated data is instructive. As all participants were randomly invited to the experiment I assume that there is no unmodelled systematic differences between players. If this assumption holds, pooling the data from different sessions yields a valid sample in order to analyze the presumably unique distribution-function, from which all players were drawn.[32]

[32] If the full player population (also B-players) were considered, then the assumption that there is no systematic difference between players due to their specific role is additionally necessary. The analysis proceeds with A-players.

By collecting the game-specific restrictions and corresponding proportion of choices that were observed in each game, over all players and over all games, I can characterize the player-type distribution $F(\alpha, \beta, \gamma)$, for which IASW preferences rationalize the data. For games that involve both, an advantageous and a disadvantageous inequality status, for an individual player within one decision, restrictions that characterize $F(\alpha, \beta, \gamma)$ involve all three parameters α, β, and γ. These restrictions are calculated in appendix AIII. (p. 35, game 8 et sqq.) Whenever a decision over payoff bundles does not (strongly) change the players' inequality status, the restriction implied by the game can -similarly to the UG procedure-[33] be graphed in a reduced two-dimensional (α, γ)- or (β, γ)-space. (figures 8)

In order to be transparent in the illustration of the model-calibration procedure, I reduce the amount of available data and exclusively focus on A-players. Of this A-player data, as is indicated by the consistency measure for A's behavior (table 3, line 3, p. 24), games 1, 13, and 31 are inconsistent with IASW. They imply A-players with negative (β, γ)-constellations. Calibration relies on the remaining 98 percent of IASW consistent A-choices. First figure 8a presents the restrictions implied by and the A-choices observed in the two-person games 3, 4, 5, 7, 21 and 27. These are decision situations in which player A only finds himself in an equal, or an advantageous inequality, position:

figure 8a about here

The arrows in figure 8a pointing to the upper halves of the game-specific threshold lines indicate the mass of the A-player type-distribution function $F(\alpha, \beta, \gamma)$, that must fall in the (β, γ)-value region associated with the respective arrow. Clearly the (β, γ)-distribution looses weight towards high (β, γ)-values, starting from 47 percent of the players, which could be purely self-interested. Also, there is evidence for incoherence in the distribution function implied by these games. While the significance of the incoherence between A5 (39 percent) and A27 (41 percent) is low and might be ignored -especially due to the particularity of the data that includes hypothetical choices- A3 (26 percent) and A4 (17 percent) are consistent between themselves and stand in contrast to the rest of the data. Indeed, the inconsistency causes problems in my attempt to identify a unique player-type distribution. Moreover, this

[33] Recall that in the stylized UG facts no large offers occur. This implies that proposers are exclusively on the advantegeous and responders exclusively on the disadvantegeous payoff side.

significant systematic shift in the player-type distribution is not a singularity when looking at the complete dataset. Incoherence of this kind is well interpreted by CR as evidence for negative reciprocity. Such interpretation is intuitive as reciprocity has the power to use the information from the sequential game structure. Thus it is likely to be an explanatory factor whenever the outcome differs systematically with a changing game history, as partially observed in the CR-data.

This notwithstanding, most small inconsistencies in the observed proportions are insignificant and thus irrelevant. When it comes to the characterization of the player-type distribution there are in principle three ways to handle the severe inconsistencies. First, having seen that IASW is a potentially successful approach, which could fit the data better than any other social preferences, one could endogenize its inequality aversion and welfare concern parameters α, β and γ, such that they become bigger, smaller and smaller, respectively after an hostile action; i.e. represent the flavor of negative reciprocity. Secondly, one could doubt that the different players in the various sessions are free from any systematic difference and repeat the analysis with subsets.[34] Thirdly, one could choose not to pay to much attention to fully rationalize a particular dataset, but focus on the idea of combining two motives in order to study their particular interaction.

If one can agree to the latter approach, IASW helps to understand more about the appropriateness of behavioral motives that compete for explanation of experimental facts. Regardless which way is ultimately best, there is something very important to learn from the CR data. While the UG did not lend itself to prove an existence of welfare concerned players, the CR-data does so. Plotting the restrictions of all the games where A is involved, in (weakly) advantageous (3, 4, 5, 7, 21, 27) as well as disadvantageous payoff-allocations (11, 18, 22, 28, 32), changes figure 8a as follows:

figure 8b about here

Combining the (α, γ)-restrictions with the implied (β, γ)-restrictions immediately shows that -however one thinks about the inconsistency problem- at least 17 percent of all A-players must have a welfare concern γ that is greater than $\frac{1}{3}$, and even up to 39 percent

[34] There are still inconsistencies in player-type distribution when looking, for instance, at the country-specific data subsets for Spain and the US.

may do so. This implication in turn is in line with the implication of the stylized UG-facts that at least 30 percent of the players are not more than moderately welfare concerned, i.e. $\gamma \leq \frac{4}{13}$ (see figure 5b). Surely such characterizations and the consistency check are only the most basic orientation in model evaluation.

6 Conclusion

A combination of two well-identified behavioral motives lead me to a model of inequality aversion and welfare concern. Introducing IASW preferences on the Ultimatum Game allowed concern for welfare to have a latent presence, but its existence could not be identified. This result is intuitive since in the UG welfare is unique and fixed, if it materializes through acceptance of an offer. When, in a next step, looking at experimental data, for which the welfare at stake could directly be influenced by players' decisions in a manner that goes beyond welfare existence, then the experimental evidence of Charness and Rabin (2002), for instance, implies that 17-39 percent of the players with IASW preferences have a substantial welfare concern with weight $\gamma > \frac{1}{3}$ in their IASW preference, in which self-interest weighs one. The welfare concern interacts with the concern for inequality aversion and trade-offs between the two motives were studied extensively. Other welfare concerned player-types than the 17-39 percent could not be depicted in the calibration but more may worry about others. In a consistency check, which was carried out on CR's set of simple decision games, IASW preferences spontaneously beat the neoclassical model, which assumes purely self-interested behavior of individuals, a model of perfect altruism, Fehr and Schmidt (1999)'s model of inequality aversion and CR's social-welfare preferences. Though remarking that the ranking according this consistency check is flawed, the ad hoc success of IASW questions the conclusion of CR (p. 849) that the "apparent adequacy of inequality aversion models has likely been an artefact of powerful and decisive confounds in the games used to construct these models."

A very instructive aspect of CR's dataset is that game history systematically matters. The interpretation of the statistical evidence as (negative) reciprocity is definitely the starting point for understanding why it is not possible to find one coherent player-type distribution for the IASW model within their data. Although this is unfortunate for the calibration of the IASW model on the CR data, it does not worry me too much because I did not intend to show that inequality aversion combined with welfare concern can completely substitute the elaborate idea of reciprocity. Reciprocity could in principle be included in the IASW model, but how to do so was not considered in this paper. The main intention was to introduce the idea of a behavioral interaction of inequality aversion and welfare concern in a parsimonious model, which competes with other hybrid-motive social preferences on an equal level. Aiming for predictive success through an extension may be valuable in further applications. This consideration notwithstanding, IASW preferences prove that inequality aversion is an important behavioral factor in an environment, which was recently claimed by a reciprocity and welfare model, once it is equally extended by welfare concern. Thereby contrasting with the findings of CR, the IASW preferences present encouraging evidence for inequality aversion.

7 Appendix

Appendix AI: *Calibration of the IASW Model - Calculating the Optimal Offer* s^*

optimal offer range			**restriction(s) stem from**		
i)	$s^* \in \left[\frac{4}{9}, 1\right]$	$\Longrightarrow$	follows directly from proposition 4 (i).		
ii)	$s^* \in \left[\frac{4}{9}, s_M\right]$	$\Longrightarrow$	follows directly from proposition 4 (ii).		
iii)	$s^* = s_M = 0.5$	$\Longrightarrow$	follows directly from proposition 4 (iii).		
iv)	$s^* = \frac{1}{3}$	$\Longrightarrow$	$E\left[u_P(\frac{1}{3})\right] > E\left[u_P(0.5)\right]$ (*)		
		and	$E\left[u_P(\frac{1}{3})\right] > E\left[u_P(\frac{4}{9})\right]$ (**)		
	recall:	$\rightarrow$	$E\left[u_P(0.5)\right] > E\left[u_P(s_H)\right]$ holds.		
		$\rightarrow$	$u_P(\frac{1}{3}) = \frac{2}{3} + \frac{1}{3}\gamma_P - \frac{1}{3}\beta_P$, $p_L(\frac{1}{3}) = 0.9$		
		$\rightarrow$	$u_P(\frac{4}{9}) = \frac{5}{9} + \frac{4}{9}\gamma_P - \frac{1}{9}\beta_P$, $p_L(\frac{4}{9}) = 1$		
		$\rightarrow$	$u_P(\frac{1}{2}) = \frac{1}{2} + \frac{1}{2}\gamma_P$, $p_L(\frac{1}{2}) = 1$		
	(*):	$p_L(\frac{1}{3}) \cdot u_P(\frac{1}{3}) > p_L(\frac{4}{9}) \cdot u_P(\frac{4}{9}) \Leftrightarrow$			
		$0.9 \cdot \left[\frac{2}{3} + \frac{1}{3}\gamma_P - \frac{1}{3}\beta_P\right] > 1 \cdot \left[\frac{5}{9} + \frac{4}{9}\gamma_P - \frac{1}{9}\beta_P\right]$		$\Leftrightarrow$	$1 - \frac{17}{4}\beta_P - \frac{13}{4}\gamma_P > 0$ (not binding)
		$p_L(\frac{1}{3}) \cdot u_P(\frac{1}{3}) > p_L(\frac{1}{2}) \cdot u_P(\frac{1}{2}) \Leftrightarrow$			
	(**)	$0.9 \cdot \left[\frac{2}{3} + \frac{1}{3}\gamma_P - \frac{1}{3}\beta_P\right] > 1 \cdot \left[\frac{1}{2} + \frac{1}{2}\gamma_P\right]$		$\Leftrightarrow$	$1 - 3\beta_P - \gamma_P > 0$ (binding)
v)	$s^* = \frac{4}{9}$	$\Longrightarrow$	$E\left[u_P(\frac{4}{9})\right] > E\left[u_P(\frac{1}{3})\right]$ (**')		
		and	$E\left[u_P(\frac{4}{9})\right] > E\left[u_P(0.5)\right]$ (***)		
	(**')	see (**) above.		$\Leftrightarrow$	$1 - \frac{17}{4}\beta_P - \frac{13}{4}\gamma_P < 0$ (binding)
	(***):	$p_L(\frac{4}{9}) \cdot u_P(\frac{4}{9}) > p_L(\frac{1}{2}) \cdot u_P(\frac{1}{2}) \Leftrightarrow$			
		$1 \cdot \left[\frac{5}{9} + \frac{4}{9}\gamma_P - \frac{1}{9}\beta_P\right] > 1 \cdot \left[\frac{1}{2} + \frac{1}{2}\gamma_P\right]$		$\Leftrightarrow$	$1 - 2\beta_P - \gamma_P > 0$ (binding)
vi)	$s^* = 0.5$	$\Longrightarrow$	$E\left[u_P(0.5)\right] > E\left[u_P(\frac{1}{3})\right]$ (*')		
		and	$E\left[u_P(0.5)\right] > E\left[u_P(\frac{4}{9})\right]$ (***')		
	(*'):	see (*) above.		$\Leftrightarrow$	$1 - 3\beta_P - \gamma_P < 0$ (not binding)
	(***')	see (***) above.		$\Leftrightarrow$	$1 - 2\beta_P - \gamma_P < 0$ (binding)

Appendix AII: *CR - Game by Game Results*

#	two-person dictator games	p^d	p^t	p^l	p^r
2	B chooses $(400, 400)$ vs. $(750, 375)$			.52	.48
17	B chooses $(400, 400)$ vs. $(750, 375)$			.50	.50
29	B chooses $(400, 400)$ vs. $(750, 400)$			.31	.69
23	B chooses $(800, 200)$ vs. $(0, 0)$			1.00	.00
8	B chooses $(300, 600)$ vs. $(700, 500)$			.67	.33
15	B chooses $(200, 700)$ vs. $(600, 600)$			.27	.73
26	B chooses $(0, 800)$ vs. $(400, 400)$			.78	.22
	two-person response games - payoffs identical				
5	A chooses $(550, 550)$ or lets B choose $(400, 400)$ vs. $(750, 400)$	.39	.61	.33	.67
7	A chooses $(750, 0)$ or lets B choose $(400, 400)$ vs. $(750, 400)$	.47	.53	.06	.94
28	A chooses $(100, 1000)$ or lets B choose $(75, 125)$ vs. $(125, 125)$	.50	.50	.34	.66
32	A chooses $(450, 900)$ or lets B choose $(200, 400)$ vs. $(400, 400)$	.85	.15	.35	.65
	two-person response games - B's sacrifice helps A				
3	A chooses $(725, 0)$ or lets B choose $(400, 400)$ vs. $(750, 375)$	.74	.26	.62	.38
4	A chooses $(800, 0)$ or lets B choose $(400, 400)$ vs. $(750, 375)$	.83	.17	.62	.38
21	A chooses $(750, 0)$ or lets B choose $(400, 400)$ vs. $(750, 375)$	.47	.53	.61	.39
6	A chooses $(750, 100)$ or lets B choose $(700, 500)$ vs. $(300, 600)$	.92	.08	.25	.75
9	A chooses $(450, 0)$ or lets B choose $(450, 350)$ vs. $(350, 450)$	.69	.31	.06	.94
25	A chooses $(450, 0)$ or lets B choose $(450, 350)$ vs. $(350, 450)$	.62	.38	.19	.81
19	A chooses $(700, 200)$ or lets B choose $(600, 600)$ vs. $(200, 700)$	.56	.44	.78	.22
14	A chooses $(800, 0)$ or lets B choose $(400, 400)$ vs. $(0, 800)$	.68	.32	.55	.45
1	A chooses $(550, 550)$ or lets B choose $(400, 400)$ vs. $(750, 375)$	.96	.04	.93	.07
13	A chooses $(550, 550)$ or lets B choose $(400, 400)$ vs. $(750, 375)$	.86	.14	.82	.18
18	A chooses $(0, 800)$ or lets B choose $(400, 400)$ vs. $(0, 800)$	.00	1.00	.56	.44
	two-person response games - B's sacrifice hurts A				
11	A chooses $(375, 1000)$ or lets B choose $(400, 400)$ vs. $(350, 350)$	.54	.46	.89	.11
22	A chooses $(375, 1000)$ or lets B choose $(400, 400)$ vs. $(250, 350)$	.39	.61	.97	.03
27	A chooses $(500, 500)$ or lets B choose $(800, 200)$ vs. $(0, 0)$	.41	.59	.91	.09
31	A chooses $(750, 750)$ or lets B choose $(800, 200)$ vs. $(0, 0)$	.73	.27	.88	.12
30	A chooses $(400, 1200)$ or lets B choose $(400, 200)$ vs. $(0, 0)$	.77	.23	.88	.12

Note: p^d, p^t, p^l and p^r abbreviate p^{end}, p^{enter}, p^{left} and p^{right}

Appendix AIII: *Analytical Examples (2, 17, 5) &*

Games that Do Not Reveal Information on Parameter Pairs (8, 6)

game		p^d	p^t	p^l	p^r
games involving only one inequality region					
(examples for an analytical solution to the two-player dictator and response games)					
2/17: B chooses $(400,400)$ vs. $(750,375)$[35]				.51	.49
$u_B(left) \gtrless u_B(right) \iff u_B(400,400) \gtrless u_B(750,375)$					
$\iff 400+400\gamma_B \gtrless 375+750\gamma_B-375\alpha_B$					
$\iff 1+15\alpha_B-14\gamma_B \gtrless 0$	$\Longrightarrow$			$>$	$<$
	(holds with *"sign"* for *"proportion"*)				
5: A chooses $(550,550)$ or lets B choose $(400,400)$ vs. $(750,400)$		.39	.61	.33	.67
$u_B(left) \gtrless u_B(right) \iff u_B(400,400) \gtrless u_B(750,400)$					
$\iff \alpha_B-\gamma_B \gtrless 0$	$\Longrightarrow$			$>$	$<$
$u_A(end) \gtrless E\left[u_A(enter)\right]$				$\Downarrow$	$\Downarrow$
$\iff u_B(400,400) \gtrless \left\{p^{left}\cdot u_A(B\ left)+p^{right}\cdot u_A(B\ right)\right\}$	A anticipates proportions				
$\iff u_B(550,550) \gtrless \{0.33\cdot u_A(400,400)+0.67\cdot u_A(750,400)\}$				$\Downarrow$	$\Downarrow$
$\iff 550+550\gamma_A \gtrless \{0.33\cdot[400+400\gamma_A]+0.67\cdot[750+400\gamma_A-350\beta_A]\}$					
$\iff 1-\frac{469}{169}\beta_A-\frac{300}{169}\gamma_A \lesseqgtr 0 \Longrightarrow$		$<$	$>$	$\Longleftarrow$	$\Longleftarrow$
games involving both inequality regions (analytical analysis for all relevant games)					
8: B chooses $(300,600)$ vs. $(700,500)$				.67	.33
$u_B(left) \gtrless u_B(right)$					
$\iff 300+600\gamma_B-300\alpha_B \gtrless 700+500\gamma_B-200\beta_B$					
$\iff 1+\frac{3}{4}\alpha_B-\frac{1}{2}\beta_B-\frac{1}{4}\gamma_B \gtrless 0$	$\Longrightarrow$			$>$	$<$
6: A chooses $(750,100)$ or lets B choose $(750,500)$ vs. $(300,600)$		.92	.08	.25	.75
$u_A(end) \gtrless E\left[u_A(enter)\right]$					
$\iff u_B(750,100) \gtrless \left\{p^{left}\cdot u_A(B\ left)+p^{right}\cdot u_A(B\ right)\right\}$					
$\iff 750+100\gamma_A-650\beta_A \gtrless \{0.25\cdot[750+500\gamma_A-250\beta_A]+0.75\cdot[300+600\gamma_A-300\alpha_A]\}$					
$\iff 1+\frac{1}{2}a_A-\frac{13}{7}\beta_A-\frac{1}{4}\gamma_A \gtrless 0$	$\Longrightarrow$	$>$	$<$		

Note: p^d, p^t, p^l and p^r abbreviate p^{end}, p^{enter}, p^{left} and p^{right}

[35] Indifference $u_B(left)=u_B(right)$ disregarded.

Appendix AIII (continued):

Games that Do Not Reveal Information on Parameter Pairs (9, 25, 19, 14, 30)

game	p^d	p^t	p^l	p^r
9: A chooses $(450, 0)$ or lets B choose $(450, 350)$ vs. $(350, 450)$	.69	.31	.06	.94
$u_A(end) \gtrless E\left[u_A(enter)\right] \Longleftrightarrow u_B(450, 0) \gtrless \left\{p^{left} \cdot u_A(B\ left) + p^{right} \cdot u_A(B\ right)\right\}$				
$\Longleftrightarrow 450 - 450\beta_A \gtrless \{0.06 \cdot [450 + 350\gamma_A - 100\beta_A] + 0.94 \cdot [350 + 450\gamma_A - 100\alpha_A]\}$				
$\Longleftrightarrow 1 + \frac{94}{95}a_A - \frac{423}{95}\beta_A - \frac{444}{95}\gamma_A \gtrless 0$ $\Longrightarrow$	$>$	$<$		
25: A chooses $(450, 0)$ or lets B choose $(450, 350)$ vs. $(350, 450)$	.62	.38	.19	.81
$u_A(end) \gtrless E\left[u_A(enter)\right] \Longleftrightarrow u_B(450, 0) \gtrless \left\{p^{left} \cdot u_A(B\ left) + p^{right} \cdot u_A(B\ right)\right\}$				
$\Longleftrightarrow 450 - 450\beta_A \gtrless \{0.19 \cdot [450 + 350\gamma_A - 100\beta_A] + 0.81 \cdot [350 + 450\gamma_A - 100\alpha_A]\}$				
$\Longleftrightarrow 1 + \alpha_A - \frac{431}{81}\beta_A - \frac{431}{81}\gamma_A \gtrless 0$ $\Longrightarrow$	$>$	$<$		
19: A chooses $(700, 200)$ or lets B choose $(600, 600)$ vs. $(200, 700)$	.56	.44	.78	.22
$u_A(end) \gtrless E\left[u_A(enter)\right]$				
$\Longleftrightarrow u_B(700, 200) \gtrless \left\{p^{left} \cdot u_A(B\ left) + p^{right} \cdot u_A(B\ right)\right\}$				
$\Longleftrightarrow 700 + 200\gamma_A - 500\beta_A \gtrless \{0.78 \cdot [600 + 600\gamma_A] + 0.22 \cdot [200 + 700\gamma_A - 500a_A]\}$				
$\Longleftrightarrow 1 + \frac{55}{94}\alpha_A - \frac{125}{47}\beta_A - \frac{211}{94}\gamma_A \gtrless 0 \Longrightarrow$	$>$	$<$		
14: A chooses $(800, 0)$ or lets B choose $(400, 400)$ vs. $(0, 800)$	.68	.32	.55	.45
$u_A(end) \gtrless E\left[u_A(enter)\right]$				
$\Longleftrightarrow u_B(800, 0) \gtrless \left\{p^{left} \cdot u_A(B\ left) + p^{right} \cdot u_A(B\ right)\right\}$				
$\Longleftrightarrow 800 - 800\beta_A \gtrless \{0.55 \cdot [400 + 400\gamma_A] + 0.45 \cdot [800\gamma_A - 800\alpha_A]\}$				
$\Longleftrightarrow 1 + \frac{18}{29}a_A - \frac{40}{29}\beta_A - \gamma_A \gtrless 0$ $\Longrightarrow$	$>$	$>$		
30: A chooses $(400, 1200)$ or lets B choose $(400, 200)$ vs. $(0, 0)$	.77	.23	.88	.12
$u_A(end) \gtrless E\left[u_A(enter)\right]$				
$\Longleftrightarrow u_B(400, 1200) \gtrless \left\{p^{left} \cdot u_A(B\ left) + p^{right} \cdot u_A(B\ right)\right\}$				
$\Longleftrightarrow 400 + 1200\gamma_A - 800\alpha_A \gtrless 0.88 \cdot [400 + 200\gamma_A - 200\beta_A] + 0.12 \cdot 0$				
$\Longleftrightarrow 1 - \frac{50}{3}\alpha_A + \frac{11}{3}\beta_A - \frac{64}{3}\gamma_A \gtrless 0$ $\Longrightarrow$	$>$	$<$		

Note: p^d, p^t, p^l and p^r abbreviate p^{end}, p^{enter}, p^{left} and p^{right}

References

[1] Andreoni, J. and J. Miller, "Giving according to GARP: An experimental study of rationality and altruism," Econometrica, 70, March 2002, pp. 737-53.

[2] Becker, G. S., "A theory of social interactions," Journal of Political Economy, 82, 1974, pp. 1063-93.

[3] Bolle, F. and A. Kritikos, "Distributional concerns: Equity- or efficiency-oriented?," Economic Letters, 73, 2001, pp.333-8.

[4] Bolton, G. E., "A comparative model of bargaining: Theory and evidence," American Economic Review, 81, 1991, pp. 1096-1136.

[5] Bolton, G. E. and A. Ockenfels, "A theory of equity, reciprocity and competition," American Economic Review, 90, 2000, pp. 166-93.

[6] Charness, G. and B. Grosskopf, "Cheap talk, information, and coordination - Experimental evidence," Economic Letters, 83, 2004, pp. 383-389.

[7] Charness, G. and M. Rabin, "Social preferences: Some simple tests and a new model," Quarterly Journal of Economics, 117, Aug. 2002, pp. 817-69.

[8] Cox, J. C., "Trust and reciprocity: Implications of game triads and social contexts," Working Paper 00-11, University of Arizona at Tuscon, 2003.

[9] Duesenberry, J., "Income, saving and the theory of consumer behavior," Harvard University Press, Cambridge, MA, 1949.

[10] Dufwenberg, M. and G. Kirchsteiger, "A theory of sequential reciprocity," Games and Economic Behavior, 47, 2004, pp. 268-98.

[11] Encyclopaedia Britannica, The New Encyclopaedia Britannica, Volume 1, 15th edition, London, 1998.

[12] Engelmann, D. and M. Strobel, "Inequality aversion, efficiency and maximin preferences in simple distribution experiments," American Economic Review, 94, 2004, pp. 857-69.

[13] Falk, A. and U. Fischbacher, "A theory of reciprocity," Games and Economic Behavior, *forthcoming*.

[14] Fehr, E. and K. M. Schmidt, "A theory of fairness, competition and co-operation," Quarterly Journal of Economics, 1999, 90, pp. 817-68.

[15] Fehr, E. and K. M. Schmidt, "Theories of fairness and reciprocity - Evidence and economic applications," in M. Dewatripont, L. Hansen and St. Turnovsky (Eds.), Advances in economics and econometrics: Theory and applications, Eighth World Congress, Volume 1, Econometric Society Monographs, Cambridge University Press, Cambridge, 2003, pp. 208-57.

[16] Geanakoplos, J., D. Pearce and E. Stacchetti, "Psychological games and sequential rationality," Games and Economic Behavior, 1, 1989, pp. 60-79.

[17] Gintis, H., "Game theory evolving," Princeton University Press, Princeton, NJ, 2000.

[18] Güth, W., H. Kliemt and A. Ockenfels, "Fairness vs. efficiency - An experimental study of mutual gift-giving", Journal of Economic Behavior and Organization, 50, 2002, pp. 465-75.

[19] Harsanyi, J. C., "Games with incomplete information played by 'bayesian' players," Management Science, 8, 1967-68, pp. 159-82, 320-34, 486-502.

[20] Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis and R. McElreath, "In search of homo economicus: Behavioral experiments in 15 small-scale societies," American Economic Review, 91, May 2001, pp. 73-78.

[21] Leibenstein, H., "Bandwagon, snob, and Veblen effects in the theory of consumers' demand," Quarterly Journal of Economics, 64, 1950, pp. 183-207.

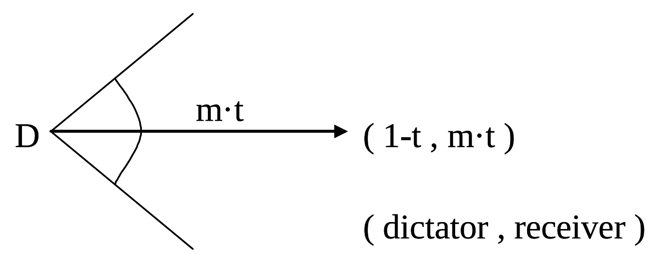
[22] Loewenstein, G., M. Bazerman and L. Thompson, "Social utility and decision making in interpersonal contexts," Journal of Personality and Social Psychology, 57, 1989, pp. 426-41.

[23] Pollak, R. A., "Interdependent preferences," American Economic Review, 66, 1976, pp. 309-20.

[24] Rabin, M., "Incorporating fairness into game theory and economics", American Economic Review, 83, 1993, pp. 1281-1302.

[25] Veblen, T., "Essays in our changing order," Edited by Leon Ardzrooni, Viking Press, New York, NY, 1934.

Figure 1: *Generalized Dictator Game*



t: transfer, m: multiplier, 1-t+mt: final surplus, m=1 standard DG, m≠1 generalized DG

Figure 2: *Ultimatum Game*

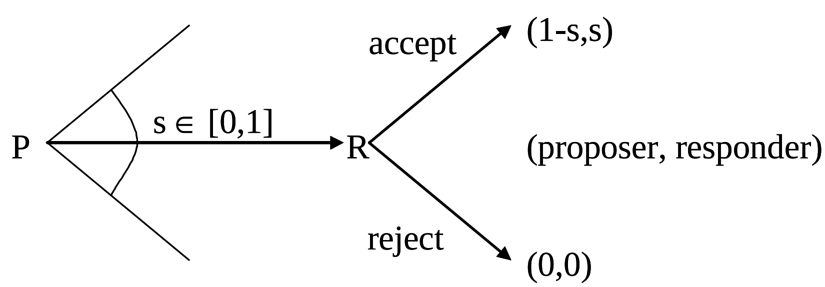


Figure 3: *Responders' critical values*

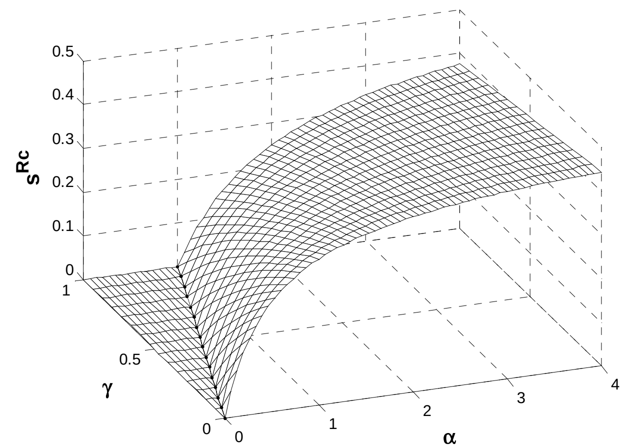
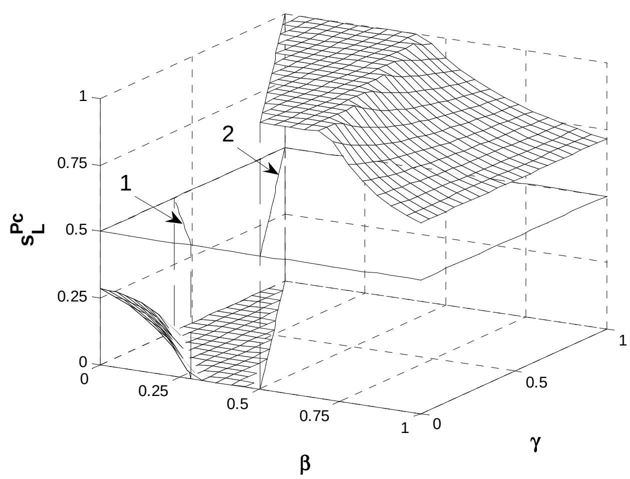
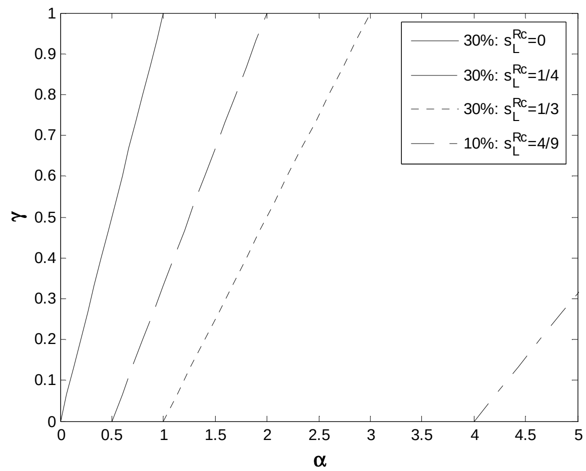


Figure 4: *Proposers' Critical Value and Choice*



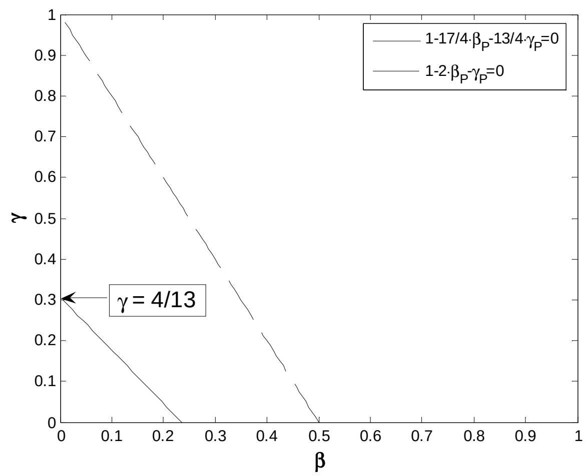
Proposer-types (β, γ) left of line 1 choose s_L^*. Types right of line 1 prefer s_M. As $p\,(s_L^*) \to 1$ line 1 converges to line 2, i.e. s_L^* is considered by more proposers.

Figure 5a: *Distribution of Acceptance Thresholds and Corresponding Responder-Types Implied by UG Data*



Responder-types, which make stylized offers, are determined by (α, γ)-values on the respective lines.

Figure 5b: *Proposer-Type Distribution Implied by UG Data*



The 30% of the proposer-types, which make offer $s^* = 1/3$ in equilibrium, have (β, γ)-values left of the solid line. Another 30%, which offer $s^* = 4/9$, have (β, γ)-values between the lines. The 40%, which offer $s = 0.5$, have (β, γ)-values right of the dashed line.

Figure 6: *Two-Person Dictator Game (left) vs.*
Two-Person Response Game (right)

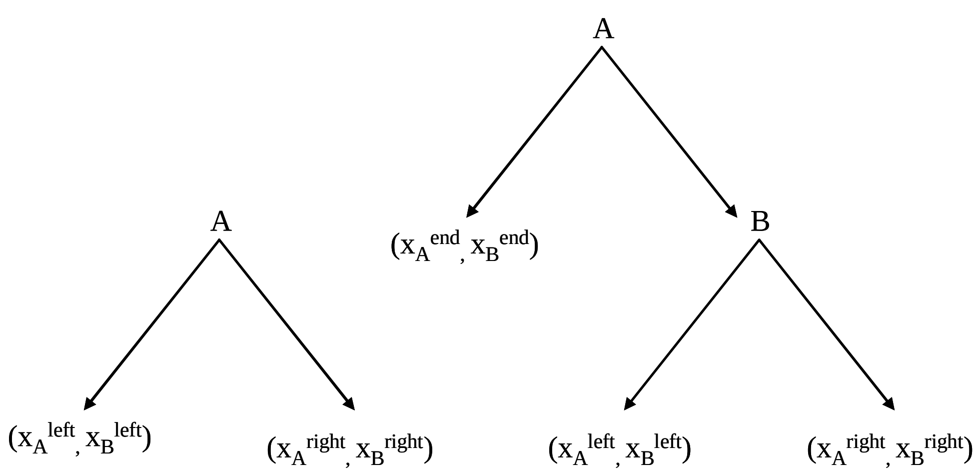


Figure 7: *Game 26 – B chooses* $(0,800)$ *vs.* $(400,400)$

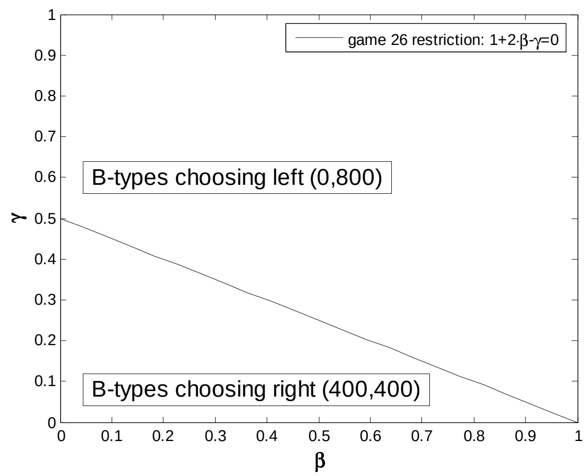


Figure 8a: *A-Players' (α, γ) Distribution*

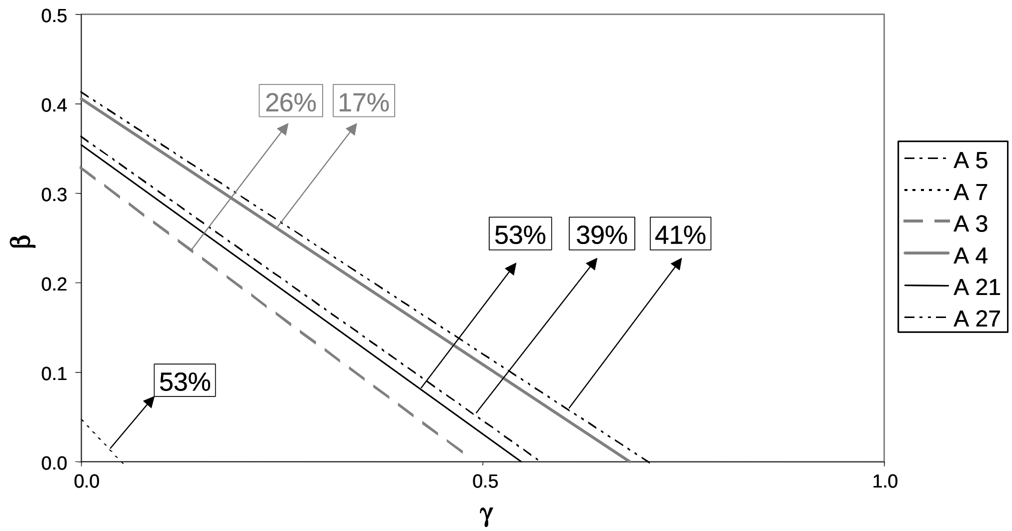


Figure 8b: *Combining A-Players' (β, γ) and (α, γ) Distribution*