Andreas Nicklisch
Irenaeus Wolff

On the Nature of Reciprocity: Evidence from the Ultimatum Reciprocity Measure

Member of

thurgauwissenschaft

www.thurgau-wissenschaft.ch

THURGAU INSTITUTE
OF ECONOMICS
at the University of Konstanz

On the Nature of Reciprocity: Evidence from the Ultimatum Reciprocity Measure*

Andreas Nicklisch

University of Hamburg &

Max Planck Institute for Research on Collective Goods, Bonn
andreas.nicklisch@wiso.uni-hamburg.de

and Irenaeus Wolff

University of Konstanz &

Thurgau Institute of Economics, Kreuzlingen
wolff@twi-kreuzlingen.ch

June 3, 2011

Abstract

We experimentally show that current models of reciprocity are in-complete in a systematic way using a new variant of the ultimatum game that provides second-movers with a marginal-cost-free punishment option. For a substantial proportion of the population, the degree of first-mover unkindness determines the severity of punishment actions even when marginal costs are absent. The proportion of these subjects strongly depends on a treatment variation: higher fixed costs of punishment lead to harsher responses. The fractions of purely selfish and inequity-averse participants are small and stable. Among the variety of reciprocity models, only one accommodates (rather than predicts) parts of our findings. The treatment effect is unaccounted for. We discuss ways of incorporating our findings into the existing models.

Keywords: *Distributional fairness, experiments, intention-based fairness, reciprocity, ultimatum bargaining*
JEL-Classification: C91, D03, D63

1   Introduction

Despite a tradition of research on reciprocal behaviour that spans almost three decades, the development of theories of reciprocal behaviour still is far from complete. One indication is that there has been a proliferation of reciprocity models (e.g., Rabin, 1993, Dufwenberg & Kirchsteiger, 2004, Sobel, 2005, Falk & Fischbacher, 2006, Cox et al., 2007) that all seem to fit specific situations better than others, and yet there is no clear indication of which model to choose in what situation. In his 2005 review article, Sobel criticizes the existing models of reciprocal behaviour for presenting a utility function of others' and own income without providing an explanation for how much weight players are likely to put on others' income relative to their own. More specifically, all of the models posit that the harshness of a reaction to an unkind action is determined by the trade-off between a reduction in the other player's payoff and the costs of punishment. For costs of punishment that are sufficiently low, these models therefore predict the harshest-possible reaction to even the slightest degree of unkindness. We argue – and show empirically – that this is wrong. However, as long as the marginal costs of punishment are strictly positive, it is impossible to falsify the above-mentioned models along these lines: it is always possible to adjust the reciprocation parameters such as to accommodate the data, given the reciprocation-parameter distribution is left unspecified in the model expositions. This substantiates a second criticism Sobel (2005, p. 407) expresses, namely that the ability of intention-based models of reciprocity to account for experimental results is "a tribute to their flexibility rather than actual support for the formulation." To corroborate the argument, we introduce the *ultimatum reciprocity measure* which eliminates the marginal costs of punishment altogether. Our experimental data show that a substantial proportion of the population deviates from the models' extreme predition in a systematic way, providing valuable insights into how existing models need to be amended.

In a recent contribution, Cox et al. (2008) abandon the domain of explicit functional forms and make a first step to address Sobel's (2005) first criticism. Our experiment suggests that their model may be an important step forward, being able to accommodate 27-47% of our observations in addition to what can be explained using the more traditional models. Nevertheless, the model still is prone to Sobel's second criticism of a lack of specificity: as we discuss in section 3, the model accommodates rather than predicts our observations. The ways in which it fails on the specificity domain will provide guidance with respect to the direction in which to refine the model.

Another question that has recently attracted attention in the scholarly discussion is that of preference heterogeneity. In the context of our game,

this particularly concerns the relative importance of intention-based reciprocal motives and inequity aversion (notably proposed by Bolton and Ockenfels, 2000, and Fehr and Schmidt, 1999). Depending on the situation, one or the other seems to dominate. In fact, there is some indication that both play a role: the results of the mini-ultimatum game experiments by Falk et al. (2003) and Cox and Deck (2005) demonstrate the importance of both approaches. When the proposer has the option to offer an equal distribution of earnings and an unequal one favoring herself, the responder rejects the latter significantly more often than when the proposer has to choose between the unequal and an even more unequal distribution of earnings (in Falk et al., 44.4% versus 8.9%). Obviously, this result points to the importance of reciprocity. However, when the proposer has no option but to choose the unequal offer, still a substantial number of responders (18%) reject. As there is no intention to favour herself on the part of the proposer, this observation suggests that inequity aversion is a second empirically relevant trigger for rejections. Other experiments have shown similar patterns (e.g., on the convex ultimatum game, Andreoni et al., 2003, on three-person ultimatum games, Bereby-Meyer & Niederle, 2005, and on a three-person gift exchange game, Thöni & Gächter, 2007).

The *ultimatum reciprocity measure* (URM game) has the following structure: a proposer makes a proposal of how to divide an endowment E.[1] The responder can either accept or reject. In the first case, the proposal is implemented, in the second, the responder obtains a fixed fraction κ of the offer x and freely chooses the proposer payoff from the interval $[0, E - \kappa x]$. The important feature of the URM game is that (in contrast to most other games with punishment in the literature) punishment is free of marginal-cost, only coming at a cost that is fixed once the offer is made.[2] This fixed cost is either equal to half the offer or to three quarters of the offer, depending on the treatment. As we will show below, models of inequity-aversion and reciprocity lead to very different predictions for behaviour in the URM game: the first class of models predicts that responders –if they reject an offer– leave the proposers with a payoff which equals their earnings. In contrast, the majority of reciprocity models predicts that responders leave the proposers with zero earnings.

The results that we obtain are striking. Less than 10% of the observations can be characterized as stemming from payoff-maximizers, models of inequity aversion account for 16-17%, traditional models of reciprocity

---

[1] A symbols table can be found in Appendix A.

[2] For games that allow for a change in the other player's payoff free of marginal costs, cf., e.g., Engelmann and Strobel, 2004, or Fisman et al., 2007, who examine this question in the dictator game.

for 18-38%.[3] At the same time, we find a substantial fraction of a fourth type that deviates from these predictions in a systematic way, which we call *gradual reciprocators*. These players are characterized by punishment patterns that leave their proposers with payoffs that are increasing in the offer made but generally lead to unequal payoffs. Moreover, the fraction of these players is determined by the treatment parameter. In the treatment with a high fixed cost of punishment, 20% of the population seem to switch from being generalized-reciprocal to conforming to traditional reciprocity models. These observations call for an extension of existing models of reciprocity in the spirit of Sobel's first criticism: a characterization of the situation that leads to the prediction of the type distribution induced by the situation.

In section 5, we discuss a number of approaches of how to modify the existing models in light of our observations. In particular, we characterize the gradual-reciprocator type within the framework of Cox et al. (2008), having dismissed the idea of matching the other's degree of kindness due to a lack of observations of the corresponding response-pattern predictions. With respect to our treatment effect, we reject an amendment of Cox et al.'s (2008) definition of *generosity* based on minimum-responder payoffs as being *ad hoc*. In contrast, we propose the situation's coerciveness, defined in terms of the gap between the player's highest and second-highest payoffs, as a promising explanation. An evaluation of the idea's predictive power, however, is beyond the scope of this article and is left for future research.

The remainder of the paper is organized as follows: section 2 introduces the URM game and presents the experimental design and procedure. Section 3 analyzes the game according to the payoff-maximization model, inequity aversion, and several types of reciprocity models, always focusing on responder behaviour. Subsequently, we analyze the experimental data with respect to these predictions and point to the existence of a player type that has received little attention in the literature so far in section 4. In section 5, we explore possible directions in which to extend existing models of reciprocal behaviour to enable them to predict the kind of behaviour observed. Finally, we summarize our findings and conclude in section 6.

---

[3]Note that we do not consider the proposers in our game; cf. section 3.

2 The game, experimental design, and procedure

2.1 The ultimatum response measure (URM game)

Like the classic ultimatum game, the URM game has two players, a proposer and a responder. The proposer is given an endowment of E and offers $x, 0 \leq x \leq E$, to the responder. If the responder accepts the offer, the proposer keeps $E - x$, while the responder earns x. If the responder rejects the offer, the responder earns κx (the *conflict payoff* π_r^c) with a commonly known parameter $\kappa \in [0, 1)$, while the proposer's conflict payoff π_p^c is any amount $y, y \in [0, E - \kappa x]$, where y is freely chosen by the responder. Therefore, the payoff functions for the proposer, π_p, and the responder, π_r, respectively, are

$$\pi_p = \begin{cases} E - x, & \text{in case of acceptance} \\ y, & \text{otherwise,} \end{cases} \quad \text{and}$$

$$\pi_r = \begin{cases} x, & \text{in case of acceptance} \\ \kappa x, & \text{otherwise.} \end{cases}$$

Figure 1 illustrates the game tree of the URM game. Note that restricting the response set to $y \equiv 0$ and setting $\kappa = 0$ yields the standard two-person ultimatum game (Güth et al., 1982).

Due to these restrictions, the standard ultimatum game provides little information about negative reciprocity as a driver for rejection (since it reduces the responder's decision to a choice between only two alternatives). In contrast, by imposing no marginal costs on responders to alter proposers' payoffs after a rejection, the 'unrestricted' URM game is able to provide a very detailed picture of subjects' motivations for rejections (as will become clear from the discussion of theoretical predictions in the next section). In particular, the lack of a trade-off between own monetary income and proposer payoff provides new insights into the nature of other-regarding preferences.

2.2 Experimental design and procedure

Each participant played one anonymous URM game either in the role of the proposer or in the role of the responder. In the instructions, we referred to proposers as *person A* and to responders as *person B*. The pie size was set to $E = 12$ euros. Offers could only be made in integers. In order to analyze individual heterogeneity of responses corresponding to different offers in greater detail, we applied the strategy-vector method to elicit responders'
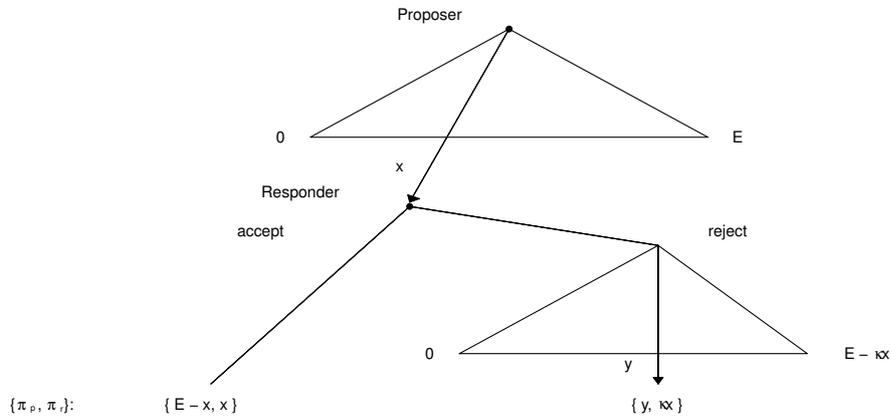
Figure 1: Game tree of the ultimatum reciprocity measure.

choices (Selten, 1967).[4] This means that responders had to make a decision for each possible (integer) offer before they were informed about the actual offer. Then, the offer and the corresponding responder decision determined the payoffs. This procedure implied that responders had to make a total of 13 acceptance/rejection decisions. Additionally, they had to determine the payoff of proposers for any offers rejected.

In contrast to the standard procedure of the strategy-vector method, responders were not provided with a choice menu, that is, a decision sheet that presents all potential offers in an ascending or descending order. Rather, potential offers were presented sequentially without a possibility of reviewing earlier decisions, and the order of possible offers differed randomly for all responders. We introduced this procedure for several reasons. The one-by-one procedure was chosen to make each decision as salient as possible. Further, eliciting decisions one by one in combination with a random order was intended to keep any potential experimenter-demand effect small by isolating decisions as much as possible: to 'smoothen' a response-pattern over all decisions out of a taste for consistency would inflict high cognitive costs

---

[4]We did not elicit hypothetical responses for accepted offers as in our view, this would have been a very counterintuitive question in the eyes of our subjects (besides the fact that it would have been difficult to incentivize such an elicitation without distorting the incentives).

on subjects. Consequently, a smooth response-pattern should only be observed if subjects exhibited underlying preferences giving rise to it. Finally, the order was randomly determined for each subject individually, in order to control for possible order effects.

The experiment started such that copies of the instructions were handed out to participants and read aloud. Subsequently, participants' questions concerning the experiments were answered privately by the instructors. Finally, all participants had to answer an electronic questionnaire testing their understanding of the game and the payoff structure.[5] Before participants answered the questionnaire, it was made clear that the only purpose of the questionnaire was to improve the understanding of the rules of the game. Wrong answers were privately explained and corrected before the experiment started.

After the elicitation of acceptance and response decisions, responders were asked to state which offer they considered as fair, and which offer they expected to receive. Subsequently, we randomly matched each responder to a proposer and payoffs were realized according to the decisions made. Participants were informed about their payoffs and asked to answer a short socio-demographic questionnaire, before privately being paid.

In order to learn more about the nature of reciprocal preferences, we played the game under two treatment conditions. In the HIGH-κ condition, the commonly known parameter κ was set to $\kappa = 0.5$, while in the LOW-κ condition, we set $\kappa = 0.25$. As we will show below, this (rather innocent) variation has little implication for the predictions of models, while there are important differences in actual behaviour. In total, 76 pairs of proposers and responders participated in the HIGH-κ treatment, while we had 77 pairs in the LOW-κ condition.

The laboratory experiments were conducted at the EconLab at the University of Bonn, Germany, in October and November 2006.[6] In total, 306 subjects participated; 50% of the subjects were female, the median age was 23 years. Participants were mostly undergraduate students from various fields of studies. Approximately one third of the students were economists or mathematicians. Further information concerning the socio-demographic background of the participants is summarized in Table C.1 in Appendix C. Average payment was 5.15 Euros (no show-up fee), corresponding to an average length of the experiment of 30 minutes, including the instruction time and the time for paying subjects.

---

[5]Translations of the German instructions and the questionnaire are provided in Appendix B.

[6]Experiments were computerized using zTree (Fischbacher, 2007). For the recruitment of subjects, we used ORSEE (Greiner, 2004).

3 Theoretical predictions

Our central research interest lies in the empirical analysis of reciprocal be-
haviour. For this reason, we will focus on the behaviour of responders
throughout the paper. Proposer behaviour is unsuitable for our purposes:
proposals reflect both proposers' other-regarding preferences as well as pro-
posers' strategic considerations concerning the other-regarding preferences of
responders.

 We will analyze responders' best-response functions according to all ma-
jor models that are potential candidates for the explanation of reciprocal
behaviour. For brevity and ease of exposition, we refrain from presenting the
complete sets of equilibria as they do not shed further light on our research
question. In the following, we discuss three (groups of) models, the 'standard'
game-theoretic prediction, models of inequity-aversion, and intention-based
models of reciprocal behaviour. Before we do so, let us clarify some notation.
If a model predicts rejection of an offer, it will have to specify a value for
the response y. This may be the case for all, some, or none of the offers,
and consequently, the response y is contingent on the offer. To reflect this,
we will write $y = y(x)$ to denote the *(offer) response function*. Yet, there is
a second way to think about responses, which will prove useful particularly
in the context of treatment comparisons. For this purpose, we introduce
the *conflict-payoff response function* (defining the response y in terms of the
conflict payoff $\pi_r^c = \kappa x$), which we will denote by $y = \gamma(\pi_r^c)$.

3.1 Pure payoff-maximizing preferences.

The best reply of a responder exclusively driven by material self-interest is
obvious: given $0 < \kappa < 1$, we have $x > \kappa x$ for any $x > 0$, and $x = \kappa x$
for $x = 0$. Consequently, payoff-maximizing responders' best reply is to
always accept any positive offer x, and arbitrarily accept or reject a proposal
of $x = 0$. Given this feature, we will not observe values of y for these
players. If at all, we observe a value for y in response to $x = 0$; however,
the theory does not give any prediction for this value. Therefore, payoff-
maximizing responders' best-response function is given by $br_{pm} : x \to (\delta, y)$,
where $\delta \in \{0, 1\}$ represents rejection, $\delta = 0$, or acceptance, $\delta = 1$:

$$br_{pm}(x) = \begin{cases} (1, .) & \text{if } x > 0, \\ (\delta', y') & \text{if } x = 0, \end{cases} \tag{1}$$

where $(\delta', y') \in \{(\delta, y) | \delta \in \{0, 1\}, y \in [0, E]\}$. Of course, no treatment differ-
ences are expected.

3.2 Inequity-averse preferences.

In a first step, note that inequity-averse responders will always choose to equalize payoffs after a rejection, since it is costless to alter the proposer's payoff once the costs of rejecting are sunk. In other words, their response y will be $y(x) = \kappa x$ for all rejected offers x. Which offers will be rejected? Both the model by Fehr and Schmidt (1999) and by Bolton and Ockenfels (2000) predict accepted offers x to come from a convex set $[\underline{x}; \overline{x}]$, where $0 \leq \underline{x} < E/2 < \overline{x} \leq E$. The specific values of $\underline{x}$ and $\overline{x}$ depend on the parameters of the model, notably on κ (since it determines the monetary earnings in case of a conflict) and the importance the individual responder places on equity concerns. To indicate the dependence between $\underline{x}$ and κ, and $\overline{x}$ and κ, we will write $\underline{x}_\kappa$ and $\overline{x}_\kappa$. Both models would suggest there to be heterogeneity in the cut-off values for rejections, while the above specification of $y(x) = \kappa x$ is unique in all models of inequity aversion. In summary, we obtain the following best-reply function $br_{ia} : x \rightarrow (\delta, y)$:

$$br_{ia}(x) = \begin{cases} (0, \kappa x) & \text{if } x > \overline{x}_\kappa \text{ or } x < \underline{x}_\kappa, \\ (1, .) & \text{if } \underline{x}_\kappa \leq x \leq \overline{x}_\kappa. \end{cases} \tag{2}$$

The predicted treatment effects are evident: an increase in κ shifts both acceptance thresholds 'inwards' towards the egalitarian payoff distribution $(E/2, E/2)$. With respect to responses as a function $\gamma(\pi_r^c)$ of conflict payoffs, no treatment differences are expected.

3.3 Intention-based preferences.

For our discussion of these models, we sub-divide this class into four sub-classes: (i) one in which utility functions consist of a linear combination of own income and a reciprocity term (which itself is a product of several terms as described below, cf. Rabin, 1993, Levine, 1998,[7] Dufwenberg & Kirchsteiger, 2004, Falk & Fischbacher, 2006), (ii) the non-linear model of Cox, Friedman, and Gjerstad (2007), (iii) models mixing reciprocity concerns and inequality aversion, and finally, (iv) the model presented by Cox, Friedman, and Sadiraj (2008).

'Linear' reciprocity models. In the models subsumed under this class, utility is a linear combination of own income and reciprocity. Here, reci-

---

[7]Strictly speaking, the model of Levine (1998) is different from the other models listed in a number of important aspects. However, the best replies are very similar, given in Levine (1998) a player's utility function is given by a linear combination of all players' monetary payoffs.

procity is a product of three terms. The first weights the importance of reciprocal behaviour to the person. The second term captures the kindness of the other player's past behaviour. Offers are ranked from unkind (i.e., small) to kind (i.e, large) ones, such that the term for an offer which is neither kind nor unkind is zero and increases (decreases) monotonically with each rank above (below) that offer. Consequently, unkind offers have negative values and kind offers have positive values. The third term measures the degree of kindness in the person's reaction. Again, responses are ranked such that a response that is neither kind nor unkind corresponds to a value of zero, and responses above (below) that lead to values increasing (decreasing) monotonically with each rank. At the time of the responder's decision in the ultimatum reciprocity measure, the first two terms are fixed. Consequently, maximization of utility in combination with the possibility of choosing the proposer's payoff free of marginal cost implies the following for rejected offers: the best reply to any unkind offer x must be the most unkind response possible, that is, $y(x) = 0$, $\forall x < \underline{x}_\kappa$. Conversely, any rejected kind offer must be answered with $y(x) = E - \kappa x$, $\forall x > \overline{x}_\kappa$, the kindest response possible. In other words, there cannot be a rejection followed by a response $y'(x_r)$, so that $0 < y'(x_r) < E - \kappa x_r$. As for the inequity-aversion models above, the switching point between acceptance and rejection is player-specific and generally cannot be predicted. The predicted best-reply function $br_{rl} : x \rightarrow (\delta, y)$ is given by:

$$br_{rl}(x) = \begin{cases} (0, E - \kappa x) & \text{if } x > \overline{x}_\kappa, \\ (1, .) & \text{if } \underline{x}_\kappa \leq x \leq \overline{x}_\kappa \\ (0, 0) & \text{if } x < \underline{x}_\kappa. \end{cases} \qquad (3)$$

Obviously, no treatment variations are predicted with respect to $y(x)$ or $\gamma(\pi_r^c)$. The lower acceptance threshold $\underline{x}_\kappa$ rises with κ, as a higher κ makes rejection less costly. At the same time, there is no clear prediction with respect to $\overline{x}_\kappa$: while a higher κ implies a higher 'conflict' payoff κx, it also leads to a lower potential for rewarding actions: $E - \kappa' x < E - \kappa'' x$ for $\kappa' > \kappa''$. Therefore, the sign of the change in the upper acceptance threshold depends on the weight the responder places on reciprocity.

Non-linear models of reciprocity. Even though Cox et al. (2007) propose a remarkable model that generalizes the above reciprocity-models in an important way, it yields the same predictions for responder behaviour in the ultimatum reciprocity measure as the 'linear' reciprocity models. In fact, utility is again a linear combination of own income and a reciprocity term, where the latter multiplies the proposer-payoff with an "emotional-state" function

θ. θ is a function of the proposer's previous behaviour, and therefore, a fixed factor at the time of the responder's decision. Consequently, the arguments from our discussion of the 'linear' models carry over and hence, the predicted best-reply function has the same form as equation (3) above.

Mixed approach   A special variation of reciprocity models is the approach by Charness and Rabin (2002) which mixes reciprocity concerns and inequality aversion.[8] In this model, a responder's utility function adds own payoff and the proposer's payoff, weighted by a term that integrates inequality as well as reciprocity concerns. In particular, this weight is lowered if the proposer receives more money than the responder and if the proposer misbehaves. However, even if the weight for reciprocity depends on the degree of misbehaviour (as in the extended model in the appendix of Charness and Rabin, 2002), the sum of weights is either positive or negative. Once again, the same arguments as for the 'linear' reciprocity models apply, leading to the same predictions.

General approach to reciprocity.   Cox et al. (2008) present their novel approach to reciprocal behaviour within the framework of the proposer-payoff–responder-payoff space. In this space, the choice set $\mathcal{S}$ of the responder consists of one point and a ray parallel to the proposer-payoff axis. The point describes the offer, while the ray characterizes possible payoff combinations in case of a rejection, as depicted in Figure 2. Notice that our treatment variation does not change the location of the point, but shifts the ray in LOW-κ closer to the proposer-payoff axis compared to the situation in HIGH-κ.

Responder preferences are represented by indifference curves $\lambda \in \Lambda$, where Λ is a player's indifference-curve set for a given situation. To illustrate, indifference curves of payoff-maximizing players are lines parallel to the proposer-payoff axis, those of inequity-averse players are either convex (Bolton & Ockenfels, 2000) or piece-wise linear with a kink at the 45-degree line (Fehr & Schmidt, 1999), indifference curves in the 'linear' reciprocity models (as well as in the 'mixed' model by Charness & Rabin, 2002) are straight lines that are either upward-sloping (negative reciprocity) or downward-sloping (positive reciprocity), while the model of Cox et al. (2007) generalizes the 'linear' reciprocity models by allowing the indifference curves to be non-linear; however, their slope cannot change signs. Irrespective of their shape, indifference curves always can be ranked such that λ' is said to be 'higher' than λ'' if points

---

[8]In fact, the approach by Falk and Fischbacher (2006) also represents a mixture of reciprocity and inequality considerations, as reciprocation by the responder is triggered by proposer choices that lead to unequal payoffs.
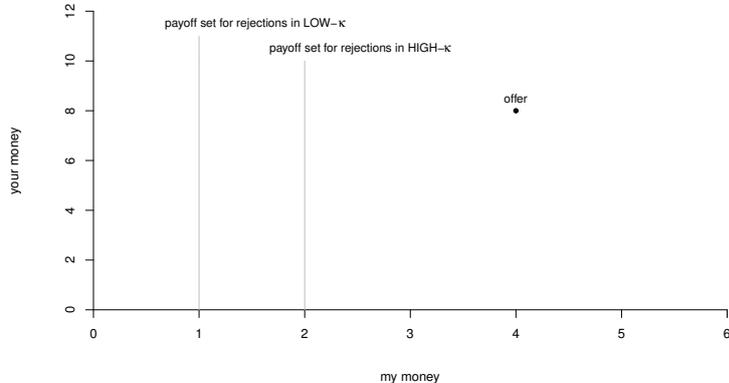
Figure 2: Payoff space for responders

associated with λ' are preferred to points associated with λ''. Finally, whenever we talk about the Λ-*defined point in* $\mathcal{S}$, we mean the point associated with the highest indifference curve in Λ which still is in the choice set $\mathcal{S}$.[9]

At the center of the approach by Cox et al. (2008) are two basic definitions, one concerns perceived kindness and the other kindness in (re-)actions. First, let us define perceived kindness, or "generosity". In this model, the notion of *generosity* is attached to responders' opportunity sets, or, more precisely, to opportunity sets after they have been altered by the action of the proposer.[10] Particularly, consider the set $\mathcal{S}_x$ of possible payoff combinations (π_p, π_r) which proposer and responder can gain after the proposer has chosen x. Let us define $\hat{\pi}_i(x) = \sup_{\pi_i} \mathcal{S}_x$ for $i = p, r$. A set $\mathcal{S}_{x'}$ is called "more generous than" a set $\mathcal{S}_{x''}$ if (i) $\hat{\pi}_r(x') - \hat{\pi}_r(x'') \geq 0$ and (ii) $\hat{\pi}_r(x') - \hat{\pi}_r(x'') \geq \hat{\pi}_p(x') - \hat{\pi}_p(x'')$. In other words, the proposer is *more generous* by choosing x' than by choosing x'' if (i) the proposer's choice of x' over x'' does not lead to a decrease in the maximum payoff the responder can earn, and (ii) the increase of responder's payoff as a result of decision x' compared to x'' is not less than the corresponding increase in the proposer's

---

[9]We do not refer to this point as the tangential point, as in case of acceptance as well as for some of the models, it would be inadequate to speak of tangents: there cannot be a tangent to a point, and in some cases, the (highest) indifference curve will have a kink at the Λ-defined point.

[10]Strictly speaking, the notion of an opportunity set as used by Cox et al. (2008) would rule out application of their model to our game, as they require opportunity sets to be convex. However, we do not see why non-convexity of opportunity sets would lead to problems in the analysis. Hence, we drop the convexity assumption, as we are convinced that their model is an important tool to understand behaviour in our game.

payoff. According to this definition, an offer x' in the URM game is *more generous than* x'' if and only if $x' \geq x''$.

Second, let us define kindness in action, which is termed "altruism". *Altruism* is attached to the responder's utility function $u_r(\pi_r, \pi_p)$ and the corresponding curvature of responder's convex indifference curves Λ in the payoff-space $\{(\pi_p, \pi_r)\}$. For convenience, it is defined in terms of a player i's willingness-to-pay for a marginal increase in the payoff of player j, $WTP_i = [\partial u_i(\pi_i, \pi_j)/\partial \pi_j]/[\partial u_i(\pi_i, \pi_j)/\partial \pi_i]$, rather than i's marginal rate of substitution $MRS_i = 1/WTP_i$. The responder's utility function $u_r'(\pi_r, \pi_p)$ is said to be "more altruistic than" $u_r''(\pi_r, \pi_p)$ if $WTP_r' \geq WTP_r'', \forall(\pi_r, \pi_p)$. Equivalently, the utility function associated with indifference curves Λ' is more altruistic than the function associated with Λ'' if, compared to Λ'', the curves in Λ' are rotated counter-clockwise (compare Figure 3). As a consequence, the proposer payoff π_p' in the Λ'-defined point in $\mathcal{S}_x$ must not be smaller than π_p'' in the Λ''-defined point in the same set.
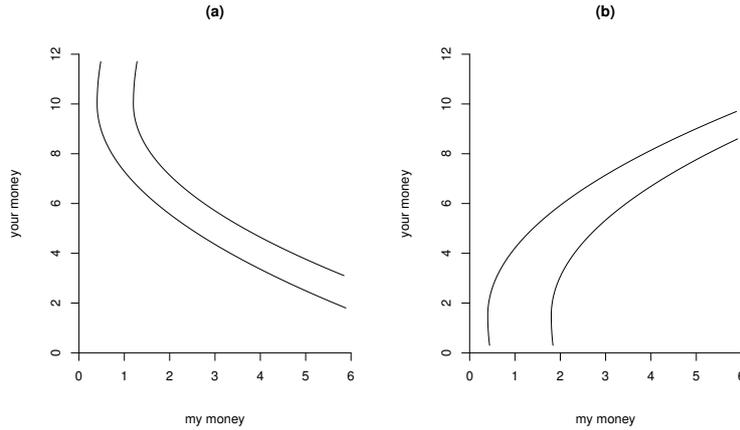
Figure 3: Indifference curves (a) Λ' and (b) Λ'' (Λ' being more altruistic than Λ'')

Within this framework, *reciprocity* is defined as follows: a proposer's decision leading to $\mathcal{S}_{x'}$ rather than $\mathcal{S}_{x''}$ ($\mathcal{S}_{x'}$ being *more generous than* $\mathcal{S}_{x''}$) induces indifference curves Λ' rather than Λ'' on the part of the responder (with $u_r'(\pi_r, \pi_p)$ *more altruistic than* $u_r''(\pi_r, \pi_p)$). Loosely speaking, more generous offers lead to more altruistic preferences.

Having outlined the model, we now apply it to the URM game. Recall that the convex indifference curves are rotated clockwise for less generous offers. That is, the smaller the offer, the steeper – or flatter, in case of upward-sloping curves – the indifference curves. As a consequence, the intersection

between the highest indifference curve and the choice set decreases or remains constant, but never increases in the proposer-payoff dimension for offers of decreasing generosity.[11] Since altering the proposer's payoff is costless and $\inf_{\pi_p} \mathcal{S}_x = 0, \forall x$ (i.e., the lower bound of the choice set does not change for different offers) we can conclude that for two rejected offers x' and x'' such that $x' > x''$, $y(x') \geq y(x'')$ must hold.

Which offers will be rejected? Like in any of the models presented, the model proposed by Cox et al. (2008) assumes that utility from own income is traded off against a second utility component that is influenced by others' income. If the responder rejects an offer, the utility gains from this second component must outweigh the decrease in one's own income. Hence, the responder must have a positive WTP in response to very *generous* offers (e.g., rejecting $x \geq \overline{x}$ and responding by $y > E - x$) – although this scenario appears hardly intuitive at the first glance – or have a negative WTP for π_p in response to very *ungenerous* offers (e.g., rejecting $x \leq \underline{x}$ and responding by $y < E - x$), so that accepted offers come from a convex set $[\underline{x}; \overline{x}]$. The specific values of $\underline{x}$ and $\overline{x}$ again depend on the importance the individual responder places on reciprocity. Thus, we obtain the following best-reply function $br_{re} : x \to (\delta, y)$:

$$br_{re}(x) = \begin{cases} (0, y'(x)) & \text{if } x > \overline{x}_\kappa \text{ or } x < \underline{x}_\kappa, \\ (1, .) & \text{if } \underline{x}_\kappa \leq x \leq \overline{x}_\kappa. \end{cases} \qquad (4)$$

where $y'(x)$ must satisfy $\partial y'/\partial x \geq 0$.

With respect to treatment effects, we first turn to changes in the offer-response function $y(x)$. In response to an increase in κ, the model allows for both monotonic increases and invariance at any given level, merely ruling out reductions in the response.[12] Turning to treatment effects on the conflict-payoff response function, recall that our treatment variation does not alter the supremum of π_r in the set $\mathcal{S}_x$ for a given offer x, since the value of κ changes the ray, but not the point (i.e., the offer, which defines the supremum). Hence, the treatment variation does not change the generosity of offers.[13] On the other hand, the same conflict payoff π_r^c is associated with different offers in the different treatments: under a low κ, a higher offer is associated with the same conflict payoff than under a high κ. At the same time, higher

---

[11]Strictly speaking, this argument requires preferences to have the *increasing benevolence* property, which Cox et al. (2008) define as a willingness to pay for the other player's income that does not decrease in own income.

[12]Once again, this requires invoking the *increasing benevolence* property, cf. ftn. 11.

[13]Strictly speaking, this statement is not correct. Please, refer to the discussion in section 5 for why we hold the above assertion to be in the spirit of the model.

offers are attached to higher levels of *generosity* and therefore, met with higher degrees of *altruism*. As a consequence, the conflict-payoff response function $\gamma(\pi_r^c)$ may differ between the treatments: for $\kappa' < \kappa''$, the same conflict payoff $\pi_r^c = \kappa' x' = \kappa'' x''$ implies $\gamma(\kappa' x') \geq \gamma(\kappa' x'')$.

With respect to acceptance thresholds, the model predicts a decrease of the acceptance threshold $\underline{x}$ for decreasing κ. To see this, take the offer x' that makes the responder indifferent between accepting and rejecting under κ'. Let us now decrease κ. Recall that a change in κ leaves the responder-payoff supremum unaffected and hence, indifference curves do not change as the responder's degree of *altruism* remains the same. But with the indifference curves remaining the same and the ray of $\mathcal{S}_x'$ shifting left, the responder must now prefer to accept the offer. By the same token, changing κ may change the upper threshold $\overline{x}$. By the convexity of preferences and the linearity with respect to variations of κ of the maximum-possible reward $(E - \kappa x)$, it is immediately obvious that an increase in κ cannot be associated with an increase in the upper acceptance threshold (and more often than not, it will lead to a decrease in $\overline{x}$). Figure 4 visualizes the effect of varying κ for the acceptance of unkind offers.
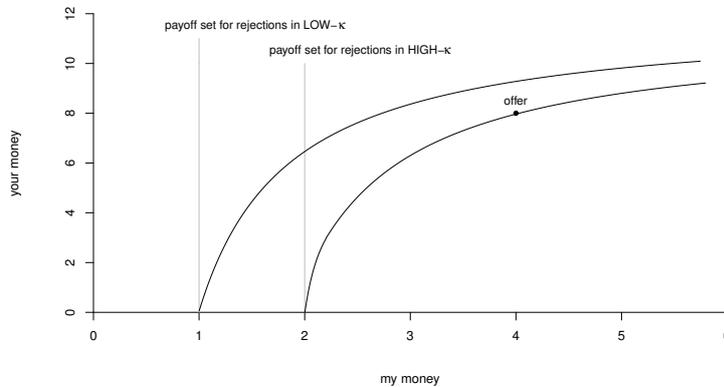
Figure 4: Example for the influence of a κ variation on offer acceptance

In the preceding paragraphs, we have presented qualitative predictions that can be derived from the general model by Cox et al. (2008). Virtually all of these predictions have been weak, in the sense exemplified by the statement that "for offers of decreasing generosity, the response cannot increase." By employing *weak* inequalities in all of their definitions, Cox et al. (2008) encompass all of the existing model predictions in one framework. The model would be able to account even for offer-response functions equal

15

to a strictly positive constant, in contrast to any of the other models. It does place a number of restrictions on behaviour that can be expected, most notably perhaps the requirement of a certain degree of consistency. However, it does not make clear predictions like the remaining models presented. In other words, what the model gains in generality, it looses in terms of specifity. We consider this an important shortcoming and briefly review potential directions of model refinement to eschew this problem in section 5 of this article. To prepare the floor for the results, we summarize the predictions from the different models in Table 1.

Table 1: *Predictions of the models discussed*

Response $y(x)$ to offers	$0 < x < \underline{x}_\kappa$	$x > \overline{x}_\kappa$
Payoff-maximization	Not applicable: all offers are accepted	
Inequity aversion	κx	$\min\{\kappa x, E - \kappa x\}$
Reciprocity (linear/non-linear/mixed)	0	$E - \kappa x$
Cox et al. (2008)	$\partial y/\partial x \geq 0$	$y > E - \overline{x}_\kappa, \partial y/\partial x \geq 0$

4 Results

We structure the presentation of our results as follows: first, we characterize rejection behaviour of responders. Second,we analyze response patterns. Third, we systematically relate response patterns to rejection behaviour. We relegate presentation of data on offers, expected offers, as well of as role-contingent average payoffs to appendix Appendix C, as these are not in the focus of this study.

4.1 Rejections

84% of actual offers in the HIGH-κ condition (81% in the LOW-κ condition) were accepted. Following our theoretical discussion from the previous section, we define an upper and a lower acceptance threshold for each responder i, $\overline{x}_i$ and $\underline{x}_i$, as follows:

$$\begin{aligned} \overline{x}_i &= max\{x|\delta_i(x) = 1\} \text{ and} \\ \underline{x}_i &= max\{x|x \leq 6, \delta_i(x) = 0\}, \end{aligned} \quad (5)$$

where $\delta_i(x)$ denotes the acceptance decision of responder i for a certain offer x. Note that inequity aversion and reciprocity predict regularity with respect

Table 2: *Numbers of responders according to acceptance thresholds*

	$\underline{x}_i$											$\overline{x}_i$
	0	1	2	3	4	5	6		9	10	11	12
HIGH-κ	3	5	5	9	42	11	1		1	1	10	64
LOW-κ	8	6	9	19	27	6	1		0	1	5	71

to rejections, that is, $\delta_i(x) = 1, \forall x \in [\underline{x}_i, \overline{x}_i]$, and $\delta_i(x) = 0$, otherwise. In total, 32 out of 153 responders exhibit rejection decisions that violate regularity. While this is more than typically observed (see Camerer, 2003), only 9 of them (6% of all responders) make more than one decision that would contradict regularity. We attribute the remaining 23 violations to the difficulty arising from the random-order one-by-one presentation of possible offers. To account for this fact and use as much information as possible, we chose the above definition of $\underline{x}_i$.[14] The further analysis includes the data of all responders. Responders are classified according to their acceptance thresholds; Table 2 reports the number of responders in each lower and upper acceptance class, $||\underline{x}_i||$ and $||\overline{x}_i||$, respectively.

The average $\underline{x}_i$ is significantly higher in HIGH-κ (3.57 vs. 2.97 in LOW-κ, $p = 0.003$; also, as can be easily seen from Table 2, $\underline{x}_i$ from the HIGH-κ treatment first-order statistically dominates $\underline{x}_i$ from the LOW-κ condition).[15] At the same time, the treatment difference between upper acceptance thresholds $\overline{x}_i$ fails to reach significance (11.80 vs. 11.91, $p = 0.114$). However, this is not a strong indication that there is no effect: while most responders never reject an offer above the equal split, the number of those who do in HIGH-κ (12 out of 76) is double the corresponding number from the LOW-κ treatment (6 out of 77; again, statistical dominance holds).

4.2 Responses

Given our main research interest lies in the study of reciprocal behaviour, responses to rejected offers are the central element in our analysis. In the following, we will identify rejected offers as x_r, so that the response to a

---

[14]Our qualitative results and statistical inferences do not change if we define $\underline{x}_i$ using the more straightforward definition $\underline{x}_i = min\{x|\delta_i(x) = 1\}$, indicating the robustness of our findings.

[15]Unless otherwise indicated, all comparisons are based on two-sided Wilcoxon rank-sum tests. Behaviour of economists/mathematicians and other participants does not differ significantly with respect to any variable measured.

rejected offer is $y(x_r)$. In light of the exploratory nature of our study, we want to get as close as possible to the raw data. Therefore, we classify the response functions according to mutually exclusive type categories. We define the types based on the theoretical predictions summarized in table 1. Participants whose responses could not be fitted into the categories defined by the presented models were grouped according to the broad characteristics of their responses.

Null. Subjects falling into this category accept all offers except for offers $x = 0$, in which case they respond by $y(0) = 0$. This behaviour can be classified as either selfish, inequity-averse or reciprocal, and therefore does not provide much information about the *nulls'* motivations.

Accepters. Subjects falling into this category accept all offers.

Symbolic. Subjects falling into this category accept most offers but reject at least one. However, their rejection is only *symbolic* – they administer the proposer the amount the latter asked for: $y(x_r) = E - x_r$.

Gentle punishers. Subjects falling into this category reject some offers but leave the proposer better off than themselves for $x_r < 6$, i.e., $y(x_r) > \kappa x_r$.

Inequity-averse. Inequity-averse players conform to the predictions of the corresponding models: whenever they reject an offer x_r which they do for at least two offers x, they respond by choosing $y(x_r) = \kappa x_r$.

Reciprocal: linear. Subjects falling into this category conform to the predictions of all major reciprocity models: whenever they reject an offer x_r which they do for at least two offers x, they respond by punishing the other player as harshly as possible, $y(x_r) = 0$.

Reciprocal: generalized. Subjects falling into this category leave proposers worse off than themselves whenever they reject an offer $x_r > 0$. However, at least for some $x_r < \underline{x}$, they return more than 0, i.e., $0 \leq y(x_r) \leq \kappa x_r$, where each inequality is strong for at least one rejected offer x_r.

Between. Subjects are categorized to fall *in between* the other categories if they reject various offers and choose $y(x_r)$ such that they would belong to different categories for different x_r.

Table 3 summarizes the results of our classification analysis. Additionally to the categories detailed above, we subdivided the category *reciprocal: generalized* into two sub-categories. This was done in order to obtain a better picture of those response patterns which have not been described in the literature up to now. The response patterns are categorized by taking the slope of a linear regression of responses $y(x_r)$ on rejected offers x_r, $x_r \leq 6$. Patterns exhibiting a slope of less than 0.2 are categorized as *static* and slopes of at least 0.2 lead to a categorization as *gradual reciprocators*. We make the following observations:

First, taking into account all response patterns that could potentially

Table 3: *Frequency of responder types*

		HIGH-κ	(in %)	LOW-κ	(in %)
Null		3	3.9	6	7.8
Accepters		1	1.3	1	1.3
Symbolic		1	1.3	1	1.3
Gentle rejecters		3	3.9	4	5.2
Inequity-averse		13	17.1	12	15.6
Reciprocal: linear		14	18.4	29	37.7
Reciprocal: generalized		36	47.4	21	27.3
	static	5	6.6	5	6.5
	gradual	31	40.8	16	20.8
Between		5	6.6	3	3.9
Total		76		77	

result from preferences of a payoff-maximizing player – *null, accepters*, and one of the *symbolic* observations – we count only 12 subjects (4 in HIGH-κ, 8 in LOW-κ) out of 153 (8%). Hence, compared to typical results from other variations of the ultimatum game (e.g., see Andreoni et al., 2003), the URM game yields much less 'selfish' behaviour by responders.

Second, the sum of all players whose behaviour can be described by one of the theoretic models outlined in section 3 excluding the model by Cox et al. (2008) makes up for only 31 out of 76 in HIGH-κ and 49 out of 77 in LOW-κ. In other words, traditional models account for only about 40% (60%) of the observed response patterns in HIGH-κ (LOW-κ). More specifically, we observe a stable 16-17% inequity-averse players, while the number belonging to different subclasses of reciprocity differs substantially across treatment conditions. Most subjects not exhibiting behaviour as predicted by the above models can be categorized as generalized-reciprocal. The model of Cox et al. (2008) can account for these observations. However, it *accommodates* rather than *predicts* them. Below, we explore a number of ways in which our understanding of generalized-reciprocal behaviour may be characterized on the basis of their model.

Third, in HIGH-κ, the fraction of players falling into the category of *gradual reciprocators* is higher than in LOW-κ by 20%. At the same time, the HIGH-κ fraction of *linear-reciprocal* players is lower by almost the same 20%. In other words, the data look as if a change in κ from 0.5 to 0.25 changed the response function of about 20% of the population such that they no longer differentiate the severity of punishment with respect to an offer's unkindness.

Table 4: *Mean lower acceptance thresholds by treatment and type*

Responder type	HIGH	LOW
Inequity-averse	4.08	3.50
Reciprocal: linear	3.71	3.31
Reciprocal: generalised		
- static	3.80	3.40
- gradual	3.97	3.00

In our view, these observations are critical: further development of any reciprocity model should account for both generalized-reciprocal behaviour and the parameter-induced shift in the type distribution, if it is to be seen as a step forward in our understanding of reciprocal behaviour. For this reason, we devote section 5 to some ideas on possible directions in which to extend existing models of reciprocity, discussing them in light of our observations. Before we do so, we shed some light on the interaction between response patterns and rejections in the following part.

4.3 The interaction between response patterns and rejections

In the following paragraph, we briefly report the results of a comparison of lower acceptance thresholds between the three main types, reported in Table 4. While we do not find any significant differences of lower acceptance thresholds between inequity-averse, linear-reciprocal, and generalised-reciprocal players within each treatment (all pair-wise comparisons yield $p > 0.15$), we observe a very differentiated picture across treatments. Both for inequity-averse and linear-reciprocal players, the treatment difference in lower acceptance thresholds is in the predicted direction but clearly fails to be significant ($p = 0.249$ and $p = 0.308$, respectively). The same holds for *static* generalised-reciprocal players ($p = 0.762$). However, for subjects classified as *gradual reciprocators*, there is a treatment effect: in LOW-κ, they accept significantly lower offers than in HIGH-κ suggesting that for this sub-class of players, fairness considerations are substantially influenced by a situational variation.

5 Generalized-reciprocal behaviour

In this section, we set out to explore possible directions in which existing models may be changed to account for our findings. Our data call for two things. The presence of a substantial fraction of subjects who can be categorized as gradual reciprocators calls for a theoretic characterization of such players. And the shift in the type distribution in response to our treatment variation calls for a theory that is able to predict that shift. Our discussion of ways to meet these challenges will be divided into two parts. First, we review and discard a simple extension of linear models of reciprocity. Subsequently, we provide a more detailed discussion within the model of Cox et al. (2008), paying tribute to the fact that it is the only available model able to accommodate our findings.

5.1 Linear reciprocity models

In the reciprocity models reviewed in this paper, there are two components of reciprocal behaviour: an assessment of the other player's kindness, or *generosity*, and the degree of reciprocation, or *altruism* in a player's response. Linear models like Charness and Rabin (2002), Dufwenberg and Kirchsteiger (2004), or Falk and Fischbacher (2006) aggregate an action's degree of *generosity* into a (relative) weight that is put on the other player's payoff; the degree of kindness of – or *altruism* in – a response then *follows from* the maximization of the weighted payoff sum. A very simple idea that would be able to meet both challenges posed by our data is to modify the models by specifying players' utility function such that it is maximized if the degree of *altruism* meets a certain target, namely the degree of *generosity* of the other player's action. This is akin to what most legal systems do: matching punishment to the severity of an offence, rather than assigning the maximum penalty to all infringements alike.

In principle, there are three ways to apply this idea to our game; however, only one of them can address both challenges posed by our data. As in the earlier reciprocity models, a responder would evaluate kindness against the 'fairness' or neutrality benchmark (in our game, presumably corresponding to the equal split) and, after a rejection, assign the proposer a fraction (i) of the offer, so that $y(x) = x/6$, (ii) of the responder's conflict payoff, yielding $y(x) = \kappa x/6$, or (iii) of the amount requested by the proposer, translating into $y(x) = (12 - x)/6$.[16] The second possibility would predict a treatment

---

[16]Another option would be to evaluate the kindness of the offer after rejection, that is, of the conflict payoff against the equal split; however, this would mean that even the equal split itself would be unkind, which is counterintuitive.

difference in the offer response function $y(x)$ as we see in our data. However, we do not want to propagate this possibility due to our empirical findings: the above conjecture (ii) makes precise point-predictions for the values of the response y that does not conform with the results for the vast majority of players categorized as gradual reciprocators even if we allow for rounding. More specifically, no point-prediction would be adequate, as the response-function variance within this group is rather large.[17] In light of this fact, it seems unsatisfactory to choose conjecture (ii) as a suitable step forward in modeling reciprocal behaviour.

5.2 Gradual reciprocity in the model of Cox et al. (2008)

Before we dwell on potential ways to account for our findings, we need to discuss an earlier imprecision in our exposition related to the reciprocity model by Cox et al. (2008).[18] We claimed that the model does not discern in terms of generosity between the same offer made in both treatments. However, blindly applying the definition of *generosity*, we would conclude that a given offer x is more generous when made in HIGH-κ than when made in LOW-κ. To see this, note that the first part of the definition, $\hat{\pi}_r^{high}(x) - \hat{\pi}_r^{low}(x) \geq 0$, trivially holds – the maximum the responder can obtain under offer x is identical in both treatments (namely, the offer itself), and thus, $\hat{\pi}_r^{high}(x) - \hat{\pi}_r^{low}(x) = 0$. At the same time, the second part also holds, as $\hat{\pi}_r^{high}(x) - \hat{\pi}_r^{low}(x) > \hat{\pi}_p^{high}(x) - \hat{\pi}_p^{low}(x)$, since for the right hand side of this equation it follows $(12 - \frac{x}{2}) - (12 - \frac{x}{4}) = -\frac{x}{4}$. In other words, the opportunity set defined by an offer x in HIGH-κ is *more generous than* the same offer in LOW-κ because it leads to the same payoff maximum for the responder but the proposer could be rewarded more in LOW-κ. This statement may be plausible when we think about very generous offers, and thus, the domain of positively reciprocal behaviour. However, for low offers x, it does not make sense conceptually: to see this, let us take the counterfactual perspective that the offer x is fixed but the proposer may choose the value of κ, and therefore, the responder's conflict payoff π_r^c. Then, application of the model definition of *generosity* would mean that a proposer would choose

---

[17] In contrast, players categorized as conforming to either inequity-aversion or reciprocity models *exactly* conform to the corresponding point-prediction (abstracting from rare acceptance-decision errors).

[18] A second imprecision is that the model is not applicable to our game if we stick to Cox et al.'s exposition, given the opportunity sets in our game are not convex, and therefore, not *opportunity sets* in the sense of their definition. However, we think it would not be conducive not to consider the model on these grounds, as it is a powerful tool to think about reciprocal behaviour, and one that deserves further development.

smaller conflict payoffs (i.e., smaller κ) because he might be *rewarded* for it – which contradicts any common notion of the concept of reciprocity. Therefore, we argue that basing the assertion that the opportunity set in HIGH-κ is *more generous than* its corresponding set in LOW-κ on the model under discussion is against its very spirit, and therefore, we contend that the model does not provide sufficient reason to predict the shift in response harshness we observe in the experiment.[19]

Having discussed Cox et al.'s definition of *generosity* at length, we are ready to review possible modifications. First of all, we suggest characterizing our gradual reciprocators by requiring the monotonicity of responses to opportunity sets – as ordered corresponding to their generosity – to be strict. In terms of the model, these players are characterized as follows: if $\mathcal{S}_{x'}$ is strictly more generous than $\mathcal{S}_{x''}$ (in the sense that the first inequality in the definition is strict), then $WTP_r' > WTP_r'' \; \forall (\pi_r, \pi_p)$. The resulting class of subjects differentiates their altruism according to the severity of an offence. On the other hand, participants with flat response patterns, most notably, linear-reciprocal subjects, no longer fall into the category.

5.3 The treatment effect on the type distribution

Our discussion from the preceding paragraphs may suggest that the critical aspect of the model on which refinement may be necessary is its definition of *generosity*. We argue that this may not be a useful approach and suggest an alternative below. Focusing on a refinement of *generosity*, a possible way capable of solving the problem in the specific context of our game would be to add a third part to the definition, comparing minimum responder payoffs in the respective opportunity sets. Then, a given offer x would lead to a higher minimum responder payoff in HIGH-κ than it would in LOW-κ, and therefore, the offer x would be seen as nicer in HIGH-κ than in LOW-κ. On the other hand, adding part after part to the definition does not seem like a sensible approach: this would extend the set of situations further in which the theory is unable to make any prediction at all.

What this discussion may suggest more broadly is that there may not be a one-size-fits-all description of *generosity*. Possibly, *generosity* has to be defined by situational classes, in conjunction with a theory that categorizes situations into these classes. Furthermore, there may be heterogeneity in people's assessment of other people's *generosity*. More radically still, it would even be conceivable that subjects differ more in their perception of *generosity* than in their reaction function. Put differently, an interesting re-

---

[19]As stressed above, this does not mean the model cannot *accommodate* the observations.

search question for future work is how much of the observed heterogeneity in behaviour is actually due to heterogeneity in perception.

How can we account for the observed type shift between treatments? We propose that the coerciveness of an offer may provide the explanation. To show this, let us suggest the following argument: a responder's position in LOW-κ seems much less comfortable than in HIGH-κ, given – holding the offer the same – the responder in LOW-κ has to renounce a larger amount of money than the responder in HIGH-κ. So, we may expect the responder in LOW-κ to be more reluctant to reject a given offer than the responder in HIGH-κ, and therefore, that the proposer's position is more powerful in LOW-κ as compared to HIGH-κ. In other words, the situation in LOW-κ can be interpreted as being more coercive; if the coerciveness of the situation determines the intensity of responders' reaction to a given offer, then we should observe the type shift we observe.[20] A possible mechanism that may give rise to the postulated effect would be that responders display an aversion to proposers abusing their power.

Is there a way in which to characterize the coerciveness of an offer? An intuitive way would be to compare the highest and the second-highest possible responder payoffs, potentially normalized using the highest-possible responder payoff. In the games examined by Cox et al. (2008), coerciveness would always be zero as they require opportunity sets to be convex; however, when applying the model to discrete opportunity sets as in our game, the hypothesis can, indeed, distinguish between different situations. To develop this idea fully and incorporate it into a modified version of the model goes beyond the scope of this article.

6   Summary and discussion

In this article, we present the ultimatum reciprocity measure (URM game) as an analytical tool for the inquiry into the nature of reciprocal behaviour. In contrast to many other games (e.g., the ultimatum game or the trust game), it gives rise to very clear and distinct predictions of models of inequity aversion on the one hand, and 'traditional' models of reciprocity, on the other. The model of Cox et al. (2008) accommodates both predictions as well as

---

[20]One may argue that proposers who make the same offer to the responder in LOW-κ than in HIGH-κ do so *despite* their more powerful position. Therefore, if proposer do not exploit their powerful position, one could conjecture that this non-exploitation is a kind act that should be rewarded under a reciprocity hypothesis. However, this is not consistent with our data: the same offer is punished as harshly as possible by a larger fraction of subjects in LOW-κ, rather than the other way around.

data that fall in between the two extremes. An important empirical aspect of our study is to provide data on the relative frequency of these types (and possibly others, if they were to be observed). Using second-movers' response patterns, we classify our subjects. Our findings are remarkable. Less than 10% of the responders in our study exhibit behaviour that can be explained by payoff-maximization; 16-17% can be classified as inequity-averse; 'traditional' models of reciprocity account for another 18-38%, depending on the treatment. This means that the main models discussed in the literature account for only 45-64% of the observations. Adding the model of Cox et al. (2008), this number increases to over 90%. We count this as evidence that the latter model is an important step forward in the quest for understanding reciprocal behaviour.

At the same time, we observe a systematic shift of behaviour between treatments that is unaccounted for by any of the models discussed in the literature. Decreasing the responder's conflict payoff by one half leads to a strong decrease in the frequency of players characterized as gradual reciprocators; at the same time, the frequency of linear-reciprocity types increases by the same amount. It seems as if the parameter difference induces one fifth of the population to respond in a *qualitatively* different way. In section 5, we discuss a number of possible ways to account for this shift. The explanation that seems to be most convincing to us – and most conducive for a further development of the theory – is that the *coerciveness* of the situation influences responders' reactions: the higher the fraction of their potential earnings they have to give up in order to be able to punish, the harsher will be their response. This could be explained if we assume that people display an aversion to the abuse of power by others.

This paper contributes to the literature in a number of important ways. It introduces the ultimatum reciprocity measure as a powerful tool that provides new insights into both the nature of reciprocal behaviour and the heterogeneity of preferences. We thereby extend the results of previous experiments that estimate interdependent preferences by using decisions in dictator games (e.g., Andreoni & Miller, 2002, Fisman et al., 2007) and other modified ultimatum games[21] which have focused predominantly on the robustness of the prediction based on inequity aversion (Kagel & Wolfe, 2001, Andreoni et al., 2003, Garrod, 2008). Moreover, the ultimatum reciprocity measure conveys valuable insights into how the heterogeneous type distribution changes as a consequence of differences in the situation, as exemplified by a simple parameter change within our game.

---

[21] For instance, see the experiments on the convex ultimatum game (e.g., see Suleiman, 1996, Charness & Rabin, 2002).

Our data provide evidence of a player type that has received little attention in the literature so far. This player type aims to level the punishment of unkind behaviour according to the degree of unkindness, rather than merely restricting their punishment in response to increasing punishment costs, as in the more traditional models of reciprocity. To the best of our knowledge, only the model proposed by Cox et al. (2008) can accommodate this behaviour. However, one might argue that the model can accommodate the behaviour because of its flexibility rather than its accuracy. We discuss one potential weakness of the model and point to a number of possible ways of how to improve on prediction accuracy, evaluating them in light of our findings. Our treatment variation seems to suggest that the coerciveness of the situation is an important determinant on the harshness of the 'new' type's responses: the more coercive the situation, the harsher the responses and the less the players bother to nuance their punishment choices. This suggests that the situation's coerciveness may be the most promising avenue in which to extend the model in future work.

References

[1] Andreoni, J., M. Castillo & R. Petrie (2003), What do bargainers' preferences look like? Experiments with a convex ultimatum game, <u>American Economic Review</u> **93**, 672-685.

[2] Andreoni, J. & J. Miller (2002), Giving according to GARP: An experimental test of the consistency of preferences for altruism, <u>Econometrica</u> **70**, 737-753.

[3] Bereby-Meyer, Y. & M. Niederle (2005), Fairness in bargaining, <u>Journal of Economic Behavior and Organization</u> **56**, 173–186.

[4] Bolton, G.E. & A. Ockenfels (2000), ERC: A theory of equity, reciprocity, and competition, <u>American Economic Review</u> **90**, 166-193.

[5] Camerer, C.F. (2003), <u>Behavioral Game Theory</u>, Princeton: Princeton University Press.

[6] Charness, G. & M. Rabin (2002), Understanding social preferences with simple tests, <u>Quarterly Journal of Economics</u> **117**, 817-869.

[7] Cox, J.C. & C.A. Deck (2005), On the nature of reciprocal motives, <u>Economic Inquiry</u> **43**, 623-635.

[8] Cox, J.C., D. Friedman & S. Gjerstad (2007), A tractable model of reciprocity and fairness, <u>Games and Economic Behavior</u> **59**, 17-45.

[9] Cox, J.C., D. Friedman & V. Sadiraj, V. (2008), Revealed Altruism, <u>Econometrica</u> **76**, 31-69.

[10] Dufwenberg, M. & G. Kirchsteiger (2004), A theory of sequential reciprocity, <u>Games and Economic Behavior</u> **47**, 268-298.

[11] Engelmann, D. & M. Strobel (2004), Inequity aversion, efficiency, and maximin preferences in simple distribution experiments, <u>American Economic Review</u> **94**, 857-869.

[12] Falk, A. & U. Fischbacher (2006), A theory of reciprocity, <u>Games and Economic Behavior</u> **54**, 293-315.

[13] Falk, A., E. Fehr & U. Fischbacher (2003), On the nature of fair behaviour, <u>Economic Inquiry</u> **41**, 20-26.

[14] Fehr, E. & K.M. Schmidt (1999), A theory of fairness, competition, and cooperation, <u>Quarterly Journal of Economics</u> **114**, 817-868.

[15] Fischbacher, U. (2007), z-Tree: Zurich toolbox for ready-made economic experiments, Experimental Economics **10**, 171-178.

[16] Fisman, R., S. Kariv & D. Markovits (2007), Individual preferences for giving, American Economic Review **97**, 1858-1876.

[17] Garrod. L. (2008), Do people really prefer equitable distributions?, working paper.

[18] Greiner, B. (2004), An online recruitment system for economic experiments. In: Kremer, K. & Macho, V. (eds.), Forschung und wissenschaftliches Rechnen 2003, GWDG Bericht 63, Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung, 79-93.

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

[20] Kagel, J. & K. Wolfe (2001), Tests of fairness models based on equity considerations in a three-person ultimatum game, Experimental Economics **4**, 203-219.

[21] Levine, D. (1998), Modeling altruism and spitefulness in experiments, Review of Economic Dynamics **1**, 593-622.

[22] Rabin, M. (1993), Incorporating fairness into game theory and economics, American Economic Review **83**, 1281-1302.

[23] Selten, R. (1967), Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In: H. Sauermann (ed.), Beiträge zur experimentellen Wirtschaftsforschung, Tübingen: J.C.B. Mohr, 136-168.

[24] Sobel, J. (2005), Interdependent preferences and reciprocity, Journal of Economic Literature **43**, 392-436.

[25] Suleiman, R. (1996), Expectations and fairness in a modified ultimatum game, Journal of Economic Psychology **17**, 531-554.

[26] Thöni, C. & S. Gächter (2007), Understanding social interaction effects in the workplace, Working Paper, University of St.Gallen.

Appendix A   Symbols table

Table A.1: *Symbols table*

Variable	Explanation
$\gamma(\pi_r^c)$	the responder's conflict-payoff response function: choice of y as a function of $\pi_r^c = \kappa x$
δ	the responder's choice variable of accepting ($\delta = 1$) or not
λ	an indifference curve
Λ	the set of (the responder's) indifference curves
κ	fraction of the proposer-offer the responder keeps after rejection
π_p	the proposer's payoff
π_r	the responder's payoff
π_r^c	the responder's (conflict) payoff in case of rejection
$\hat{\pi}_i^{(trmt)}(x)$	i's supremum payoff in the responder's choice set $\mathcal{S}_x$ (in treatment *trmt*)
br_j	a j-type's best-response function
E	a pair's endowment
$u_r(\pi_r, \pi_p)$	the responder's utility function
WTP_i	i's willingness to pay for a marginal increase in j's payoff
x	the proposer's offer
$\underline{x}_{(\kappa)}$	the lower acceptance threshold
$\overline{x}_{(\kappa)}$	the upper acceptance threshold
y	proposer payoff after rejection as determined by the responder
$y(x)$	the responder's offer response function
$\mathcal{S}_{(x)}$	the responder's choice set in the proposer-payoff–responder-payoff space (as determined by x)

Appendix B   Instructions and Questionnaire

Instructions[22]

Thank you for participating in our experiment! In this experiment, you will make decisions with which you can earn money. How much you will earn depends on your decisions and the decisions of other participants. Please

---

[22]The following instructions and the questionnaire are translations of the German originals that are available from the authors upon request. Treatment variations are indicated by brackets.

read these instructions very carefully. We kindly ask you to refrain from any public announcements and attempts to communicate directly with other participants during the experiment. If you violate this rule, we have to exclude you from this experiment. If you have any questions, please contact one of the persons running the experiment; he or she will come to your place and clarify your questions.

The decision task

In this experiment, you and another person will interact only once. Except for us, the experimenters, no participant is able to identify the other participant. At the beginning of the experiment, one of the two persons will be randomly selected as person A, the other person as person B. Person A receives an amount of 12 euros from the experimenter. Out of these 12 euros, she offers a proposal to person B. The proposal can be any amount between 0 and 12 euros. However, the proposal must be an amount in whole euros. If person B accepts the proposal, she earns the proposal. Person A earns the rest of the 12 euros (12 - proposal). If B rejects the proposal, she earns half [a quarter] of the proposal and decides on the amount person A will earn. This can be any amount from the remaining rest (between 0 and $12 - \text{proposal}/2$) [(between 0 and $12 - \text{proposal}/4$)].

Example: Person A proposes 5 euros. If person B accepts the proposal, person A earns 7 euros $(12 - 5)$. If person B rejects the proposal, person B earns 2.50 euros (proposal/2) [1.25 euros (proposal/4)]. Then, person B determines the earnings of person A, choosing an amount between 0 and 9.50 $(12 - 2.50)$ [0 and 10.75 $(12 - 1.25)$] euros.

The setup of the experiment

Before we start with the experiment, you will receive an electronic questionnaire which you have to fill in completely. The questionnaire helps you to understand the rules of the experiment. After all participants have completed the questionnaire correctly, you will be randomly selected as person A or person B. Then, person A has to decide on the proposal she will offer to person B. The proposal must be an amount in whole euros. At the same time, person B has to decide whether to accept or to reject for all possible proposals (0 euros, 1 euro,..., 12 euros), and, if rejecting an offer, which amount person A will earn. The possible proposals will be shown to person B in random order. Please note that you cannot change your decision once you have confirmed it. Thereafter, we will ask person B which proposal

she considers as fair and which proposal she expects to receive from person
A. When each person B has decided on all possible proposals, we will select
randomly and anonymously for each person A a person B. Payoffs for person
A and person B are determined by the proposal of person A and person B's
decision, which she determined for the particular proposal beforehand. This
means, that each decision of person B is relevant for the determination of the
payoffs. At the end of the experiment, we will inform you of your payoff and
ask you to answer a short socio-demographic questionnaire (your age, your
sex, etc.). Then you will receive your payoff privately; no other participant
will be able to see what you have earned.

Questionnaire

Please mark the correct answers for the following questions. They will help
you to understand the rules of the experiment. Please note that there could
be more than one correct answer, so that you have to mark in those cases
more than one answer.

(1) Suppose person A offers 6 euros. Person B rejects. How much does
person B earn?

    a. 1 [0.50] euros

    b. 3 [1.50] euros

    c. 6 euros

(2) In the situation described in question (1), from which interval does
person B choose person A's payoff?

    a. 0 to 12 euros

    b. 0 to 6 euros

    c. 0 to 9 [10.5] euros

(3) Suppose person A offers 11 euros. Person B accepts. How much does
person A earn?

    a. 1 euro

    b. 5.50 euros

    c. 11 euros

(4) Suppose person B receives a proposal of 0 euros and rejects the pro-
posal. Which statement is correct?

a. Person B earns 0 euros.

b. Person B earns 0.50 euros.

c. Person B chooses person A's earnings from the interval 0 to 6 euros.

d. Person B chooses person A's earnings from the interval 0 to 12 euros.

Appendix C Participants' socio-economic background & offer and payoff data

Table C.1: *Socio-demographic summary of participants*

Age	mean: 23.8 (4.1)	range: 18-57
Number of siblings	mean: 1.47 (1.23)	range: 0-8
Female	50.0%	
Work experience†	26.9%	
Major‡	0: 7.5%, 1: 3.3%, 2: 19.3%, 3: 40.3%, 4: 29.5%	
Experimental experience*	1: 11.5%, 2: 28.5%, 3: 24.9%, 4: 35.1%	

Note: Standard deviations in parentheses; †the question was: "Do you work for 10 or more hours a week?"; ‡categories: 0: no student, 1: medicine, 2: human science, 3: natural science, 4: mathematics and economics; *categories: 1: novice, 2: less than 3 experiments, 3: less than 6 experiments, 4: more than 5 experiments.

Table C.2: *Average payoffs and offers*

	HIGH-κ	LOW-κ
Proposer payoff	5.52 (1.90)	5.43 (2.61)
Responder payoff	5.24 (1.59)	4.42 (2.01)
Offers**	5.57 (1.14)	4.91 (1.43)
Expected offers	5.55 (1.51)	5.37 (1.50)
Stated 'fair' offers	6.03 (0.89)	5.94 (0.95)

Note: Treatment averages, standard deviations in parentheses; treatment difference significant *at $\alpha = 0.05$, **at $\alpha = 0.01$, ***at $\alpha = 0.001$.

THURGAU INSTITUTE
OF ECONOMICS
at the University of Konstanz

Hauptstr. 90
CH-8280 Kreuzlingen 2

Telefon: +41 (0)71 677 05 10
Telefax: +41 (0)71 677 05 11

info@twi-kreuzlingen.ch
www.twi-kreuzlingen.ch