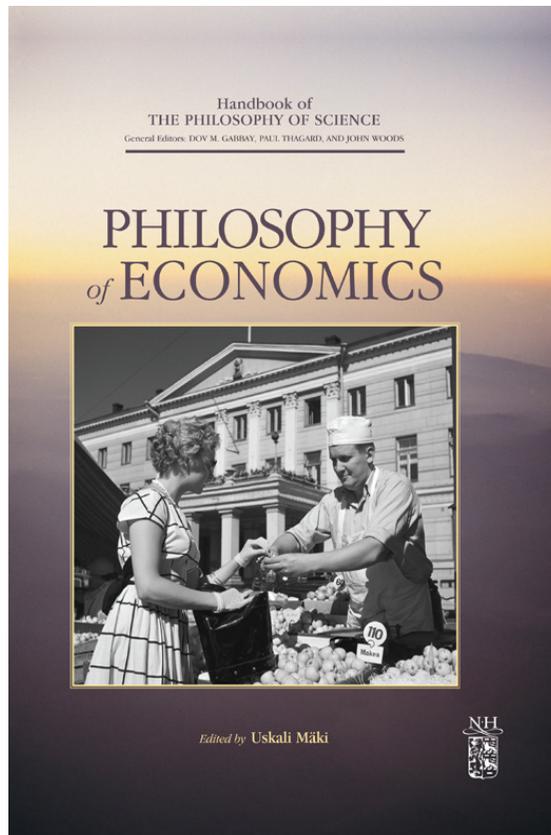


**Provided for non-commercial research and educational use only. Not for reproduction, distribution or commercial use.**

This chapter was originally published in the book *Philosophy of Economics*. The copy attached is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research, and educational use. This includes without limitation use in instruction at your institution, distribution to specific colleagues, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at: <http://www.elsevier.com/locate/permissionusematerial>

From Cristina Bicchieri and Jiji Zhang, An Embarrassment of Riches: Modeling Social Preferences in Ultimatum Games. In: Uskali Mäki, editor, *Philosophy of Economics*. San Diego: North Holland, 2012, pp. 577-595.

ISBN: 978-0-444-51676-3

© Copyright 2012 Elsevier B. V.

North Holland.

# AN EMBARRASSMENT OF RICHES: MODELING SOCIAL PREFERENCES IN ULTIMATUM GAMES

Cristina Bicchieri and Jiji Zhang

Traditional economic models presume that individuals do not take an interest in the interests of those with whom they interact. More particularly, the assumption of *non-tuism* implies that the utility function of each individual, as a measure of her preferences, is strictly independent of the utility functions of those with whom she interacts. Philip Wicksteed introduced the concept of non-tuism, stressing that an economic relationship is one entered into by two parties each of whom is intent on the furtherance of his own (not necessarily selfish) purposes, not those of the other.<sup>1</sup> Interestingly, this idea is quite different from the usual egoistic assumption: a non-tuist may be a caring, altruistic human being, but when involved in an economic exchange, she must necessarily regard her own interest as paramount. Thus non-tuism is important insofar as it defines the scope of economic activities. When tuism to some degree motivates one's conduct, then it ceases to be wholly economic.

There is nothing wrong in saying that exchange activities display the above kind of motivation, but it is certainly farfetched to assume that *all* activities we may model within a rational choice framework share the same, non-tuistic motivation. In fact, it is plainly untrue. Note that we are distinguishing here between non-tuism and selfishness, the latter being a more encompassing disposition that is not particularly tied with specific activities. That is, a selfish person will display a 'me first' attitude in all sorts of environments, caring just about her material well-being to the exclusion of other motives, whereas non-tuism is appropriate in all those cases in which we are expected to 'win' (as in competitive games or market interactions). Non-tuism, however, may not be appropriate at all in personal exchanges, and even in traditional economic areas such as labor economics [Fehr

---

<sup>1</sup>"He is exactly in the position of a man who is playing a game of chess or cricket. He is considering nothing except his game. It would be absurd to call a man selfish for protecting his king in a game of chess, or to say that he was actuated by purely egoistic motives in so doing. It would be equally absurd to call a cricketer selfish for protecting his wicket, or to say that in making runs he was actuated by egoistic motives qualified by a secondary concern for his eleven. The fact is that he has no conscious motive whatever, and is wholly intent on the complex feat of taking the ball. If you want to know whether he is selfish or unselfish you must consider the whole organization of his life, the place which chess-playing or cricket takes in it, and the alternatives which they open or close. At the moment the categories of egoism and altruism are irrelevant." (P. Wicksteed, *Common Sense of Political Economy*, p. 175.)

*et al.*, 1998] and public choice we know that cooperation, trust and fairness play a crucial role in smoothing interactions and producing better collective outcomes.

Experimental results in Ultimatum, Trust and Social Dilemma games have been interpreted as showing that individuals are, by and large, not driven by selfish motives. But we do not need experiments to know that. In our view, what the experiments show is that the typical economic auxiliary hypothesis of non-tuism should not be generalized to other contexts. Indeed, we know that when the experimental situation is framed as a market interaction, participants will be more inclined to keep more money, share less, and disregard other participants' welfare [Hoffman *et al.*, 1994]. When the same game is framed as a fair division one, participants overall show a much greater concern for the other parties' interests. The data thus indicate that the *context* of an interaction is of paramount importance in eliciting different motives. The challenge then is to model utility functions that are general enough to subsume a variety of motives and specific enough to allow for meaningful, interesting predictions to be made.

For the sake of simplicity (and brevity), in what follows we will concentrate upon the results of experiments that show what appears to be individuals' disposition to behave in a fair manner in a variety of circumstances [Camerer, 2003], though what we are saying can be easily applied to other research areas. Such experimental results have been variously interpreted, each interpretation being accompanied by a specific utility function. We shall consider three such functions and the underlying interpretations that support them, and assess each one on the basis of what they claim to be able to explain and predict.

## 1 SOCIAL PREFERENCES

The experiments we are going to discuss are designed in the form of strategic games, among which the Ultimatum game is one of the simplest [Guth *et al.*, 1982]. The game involves two players, a proposer and a responder. In the typical setting, the proposer is given a fixed amount of money and must decide how much to offer to the responder. The responder either accepts the offer, in which case the two players get their respective share, or rejects the offer, in which case both players get nothing. We will refer to this basic Ultimatum game as BUG henceforth. If, as usually assumed, the players simply aim at maximizing their monetary payoffs, the responder will accept as long as the offer is positive, and hence the proposer will offer the least amount allowed.<sup>2</sup> This prediction, however, is not confirmed by the experimental results: the modal offer is about half of the stake, and relatively mean offers are frequently rejected [Camerer, 2003]. This suggests that players' utilities are not simply increasing functions of the monetary gains.

A few utility functions have thus been developed that explicitly incorporate concerns for fairness into the preference structure. In this paper we compare

---

<sup>2</sup>We assume, throughout this paper, that the utility functions of the players are common knowledge, though particular parameter values in the functions may not be.

three such models in the context of ultimatum games — both the BUG and some variants. The three models represent very different approaches. One common explanation of the experimental data is that individuals have a *preference* for fairness in the outcome. That is, given two outcomes, individuals by and large will prefer the fairest one. The Fehr-Schmidt model is the best known example of such an approach: it is a consequentialist model, in which an agent's utility is completely determined by the final distribution of outcomes — the agent's own material payoff and others' material payoffs. By contrast, the model developed by Rabin emphasizes the role of actual actions and beliefs in determining utility. Players' utilities are not determined solely by the final distribution, but are also affected by how that distribution comes into being. This approach highlights the fact that people care about the process through which an outcome occurs, as well as the intentions of decision makers. Thus individuals will be nice to those they perceive as nice, and mean to those they believe to be mean. The alternative explanation we propose is that, in the right kind of circumstances, individuals *obey fairness norms*. To say that we obey fairness norms differs from assuming that we have a 'preference' for fairness. To conform to a fairness norm, we must have the right kind of expectations [Bicchieri, 2000; 2006]. That is, we must expect others to follow the norm, too, and also believe that there is a generalized expectation that we will obey the norm in the present circumstances. The preference to obey a norm is *conditional* upon such expectations.<sup>3</sup> Take away some of the expectations, and behavior will significantly change. A *conditional preference* will thus be stable under certain conditions, but a change in the relevant conditions may induce a predictable preference shift. The Bicchieri model, to a certain extent, combines the two formerly discussed approaches. On the surface it is very similar to the Fehr-Schmidt model; however, the distinguishing feature of the model — the incorporation of beliefs about (social) norms into the utility function — is closer in spirit to the Rabin model.

A norm-based theory makes testable predictions that are quite different, at least in some critical instances, from the predictions of theories that postulate a social preference for fairness. Theories that stress the importance of intentions and process over outcomes are more compatible with a norm-based scenario though, as we shall see in the next sections, the latter has greater predictive power. All the above theories, it must be noted, assume that individuals have social preferences, in that they take into consideration others' utilities when they make a choice. *How* this is done is a matter of contention, and is precisely what differentiates the three models we present.

## 2 ANALYSIS BASED ON BUG

In this section we introduce the three models and apply them to the basic ultimatum game (BUG), where the total amount of money to be divided is denoted by

---

<sup>3</sup>The conditions for following a norm are formally described in Chapter 1 of Bicchieri [2006].

$M$ . The proposer will offer the responder an amount  $x$  between 0 and  $M$ . If the responder accepts the proposal, the proposer gets  $M - x$ , and the responder gets  $x$ . Otherwise both get zero.

### 2.1 Fairness preferences: the Fehr-Schmidt model

The first model we consider was proposed by Fehr and Schmidt [1999]. Such model intends to capture the idea that people dislike, to a certain extent, unequal outcomes, even if they benefit from the unequal distribution. Given a group of  $L$  agents, the Fehr-Schmidt utility function of agent  $i$  is

$$U_i(x_1, \dots, x_L) = x_i - \frac{\alpha_i}{L-1} \sum_j \max(x_j - x_i, 0) - \frac{\beta_i}{L-1} \sum_j \max(x_i - x_j, 0)$$

where  $x_j$  denotes the material payoff agent  $j$  gets. One constraint on the parameters is that  $0 < \beta_i < \alpha_i$ , which indicates that people dislike inequality against them more than they do inequality favoring them. We may think of  $\alpha$  as an ‘envy’ weight, and  $\beta$  as a ‘guilt’ weight. The other constraint is  $\beta_i < 1$ , meaning that an agent does not suffer terrible guilt when she is in a relatively good position. For example, a player would prefer getting more without affecting other people’s payoff even though that results in an increase of the inequality.

Applying the model to BUG, the utility function is simplified to

$$U_i(x_1, x_2) = x_i - \begin{cases} \alpha_i(x_{3-i} - x_i) & \text{if } x_{3-i} \geq x_i \\ \beta_i(x_i - x_{3-i}) & \text{if } x_{3-i} < x_i \end{cases} \quad i = 1, 2$$

Obviously if the responder rejects the offer, both utility functions are equal to zero, that is,  $U_{1reject} = U_{2reject} = 0$ . If the responder accepts an offer of  $x$ , the utility functions are as follows:

$$U_{1accept}(x) = \begin{cases} (1 + \alpha_1)M - (1 + 2\alpha_1)x & \text{if } x \geq M/2 \\ (1 - \beta_1)M - (1 - 2\beta_1)x & \text{if } x < M/2 \end{cases}$$

$$U_{2accept}(x) = \begin{cases} (1 + 2\alpha_2)x - \alpha_2M & \text{if } x < M/2 \\ (1 - 2\beta_2)x + \beta_2M & \text{if } x \geq M/2 \end{cases}$$

The responder should accept the offer if  $U_{2accept}(x) > U_{2reject} = 0$ . Solving for  $x$  we get the *threshold for acceptance*:  $x > \alpha_2 M / (1 + 2\alpha_2)$ . Evidently if  $\alpha_2$  is close to zero – which indicates that player 2 (the responder) does not care much about being treated unfairly – the responder will accept very mean offers. On the other hand, if  $\alpha_2$  is sufficiently large, the offer has to be close to a half to be accepted. In any event, the threshold is not higher than  $M/2$ , which means that hyper-fair offers (more than half) are not necessary for the sake of acceptance.

For the proposer, the utility function is monotonically decreasing in  $x$  when  $x \geq M/2$ . Hence a rational proposer will not offer more than half of the cake. Suppose  $x \leq M/2$ ; two cases are possible depending on the value of  $\beta_1$ . If  $\beta_1 > 1/2$ ,

that is, if the proposer feels sufficiently guilty about treating others unfairly, the utility is monotonically increasing in  $x$ , and his best choice is to offer  $M/2$ . On the other hand, if  $\beta_1 < 1/2$ , the utility is monotonically decreasing in  $x$ , and hence the best offer for the proposer is the minimum one that would be accepted, i.e. (a little bit more than)  $\alpha_2 M / (1 + 2\alpha_2)$ . Finally, if  $\beta_1 = 1/2$ , it does not matter how much the proposer offers, as long as it is between  $\alpha_2 M / (1 + 2\alpha_2)$  and  $M/2$ . Note that the other two parameters,  $\alpha_1$  and  $\beta_2$ , are not identifiable in Ultimatum games.

As noted by Fehr and Schmidt, the model allows for the fact that individuals are heterogeneous. Different  $\alpha$ s and  $\beta$ s correspond to different types of people. Although the utility functions are common knowledge, the exact values of the parameters are not. The proposer, in most cases, is not sure what type of responder she is facing. Along the Bayesian line, her belief about the type of the responder can be formally represented by a probability distribution  $P$  on  $\alpha_2$  and  $\beta_2$ . When  $\beta_1 > 1/2$ , the proposer's rational choice does not depend on what  $P$  is. When  $\beta_1 < 1/2$ , however, the proposer will seek to maximize the expected utility:

$$EU(x) = P(\alpha_2 M / (1 + 2\alpha_2) < x) \times ((1 - \beta_1)M - (1 - 2\beta_1)x)$$

Therefore, the behavior of a rational proposer in UG depends on her own type ( $\beta_1$ ) and her belief about the type of the responder. The experimental data suggest that for many proposers, either  $\beta$  is large ( $\beta > 1/2$ ) or they estimate the responder's  $\alpha$  to be large. On the other side, the choice of the responder depends on his type ( $\alpha_2$ ) and the amount of the offer.

The apparent advantages of the Fehr-Schmidt utility function are that it can rationalize both positive and negative outcomes, and that it can explain the observed variability in outcomes with heterogeneous types. One of the major weaknesses of this model, however, is that it has a consequentialist bias: players only care about final distributions of outcomes, not about how such distributions come about.<sup>4</sup> As we shall see, more recent experiments have established that how a situation is framed matters to an evaluation of outcomes, and that the same distribution can be accepted or rejected depending on 'irrelevant' information about the players or the circumstances of play. Another difficulty with this approach is that, if we assume the distribution of types to be constant in a given population, then we should observe, overall, the same proportion of 'fair' outcomes in Ultimatum games. Not only does this not happen, but we also observe individual inconsistencies in behavior across different situations in which the monetary outcomes are the same. If we assume, as is usually done in economics, that individual preferences are stable, then we would expect similar behaviors across Ultimatum games. If instead we conclude that preferences are context-dependent, then we should provide a mapping from contexts to preferences that indicates in a fairly predictable way how

---

<sup>4</sup>This is a *separability* of utility assumption: what matters to a player in a game is her payoff at a terminal node. The way in which that node was reached, and the possible alternative paths that were not taken are irrelevant to an assessment of her utility at that node. Utilities of terminal node payoffs are thus separable from the path through the tree, and from payoffs on unchosen branches.

and why a given context or situation changes one's preferences. Of course, different situations may change a player's expectation about another player's envy or guilt parameters, and we could thus explain why a player may change her behavior depending upon how the situation is framed. In the case of Fehr and Schmidt's utility function, however, experimental evidence that we shall discuss later implies that a player's *own*  $\beta$  (or  $\alpha$ ) changes value in different situations. Yet nothing in their theory explains why one would feel consistently more or less guilty (or envious) depending on the decision context.

## 2.2 A conditional preference for following norms: the Bicchieri model

The norm-based utility function introduced by Bicchieri [2006] tries to capture the idea that, when a social norm exists, individuals will show different 'sensitivities' to it, and this should be reflected in their utility functions. Consider a typical  $n$ -person (normal-form) game. For the sake of formal treatment, we represent a norm as a (partial) function that maps what the player expects other players to do into what the player "ought" to do. In other words, a norm regulates behavior conditional on other people's (expected) behaviors. Denote the strategy set of player  $i$  by  $S_i$ , and let  $\mathbf{S}_{-i} = \prod_{j \neq i} S_j$  be the set of strategy profiles of players other than  $i$ . Then a norm for player  $i$  is formally represented by a function  $N_i: L_{-i} \rightarrow S_i$ , where  $L_{-i} \subseteq \mathbf{S}_{-i}$ . Two points are worth noting. First, given the other players' strategies, there may or may not be a norm that prescribes how player  $i$  ought to behave. So  $L_{-i}$  need not be — and usually is not — equal to  $\mathbf{S}_{-i}$ . In particular,  $L_{-i}$  could be empty in the situation where there is no norm whatsoever to regulate player  $i$ 's behavior. Second, there could be norms that regulate joint behaviors. A norm, for example, that regulates the joint behaviors of players  $i$  and  $j$  may be represented by  $N_{i,j}: L_{-i,-j} \rightarrow S_i \times S_j$ . Since we are concerned with a two-person game here, we will not further complicate the model in that direction.

A strategy profile  $s = (s_1, \dots, s_n)$  *instantiates* a norm for  $j$  if  $s_{-j} \in L_{-j}$ , that is, if  $N_j$  is defined at  $s_{-j}$ . It *violates* a norm if for some  $j$ , it instantiates a norm for  $j$  but  $s_j \neq N_j(s_{-j})$ . Let  $\pi_i$  be the payoff function for player  $i$ . The norm-based utility function of player  $i$  depends on the strategy profile  $s$ , and is given by

$$U_i(s) = \pi_i(s) - k_i \max_{s_{-j} \in L_{-j}} \max_{m \neq j} \{\pi_m(s_{-j}, N_j(s_{-j})) - \pi_m(s), 0\}$$

$k_i \geq 0$  is a constant representing  $i$ 's sensitivity to the relevant norm. Such sensitivity may vary with different norms; for example, a person may be very sensitive to equality and much less so to equity considerations. The first maximum operator takes care of the possibility that the norm instantiation (and violation) might be ambiguous in the sense that a strategy profile instantiates a norm for several players simultaneously. We conjecture, however, that this situation is rare, and under most situations the first maximum operator degenerates. The second maximum operator ranges over all the players other than the norm violator. In plain words, the discounting term (multiplied by  $k_i$ ) is the maximum payoff deduction resulting from all norm violations.

In the Ultimatum game, the norm we shall consider is the norm that prescribes a fair amount the proposer ought to offer. To represent it we take the norm functions to be the following: the norm function for the proposer,  $N_1$ , is a constant  $N$  function, and the norm function for the responder,  $N_2$ , is nowhere defined.<sup>5</sup> If the responder (player 2) rejects, the utilities of both players are zero.<sup>6</sup>

$$U_{1reject}(x) = U_{2reject}(x) = 0$$

Given that the proposer (player 1) offers  $x$  and the responder accepts, the utilities are the following:

$$\begin{aligned} U_{1accept}(x) &= M - x - k_1 \max(N - x, 0) \\ U_{2accept}(x) &= x - k_2 \max(N - x, 0) \end{aligned}$$

where  $N$  denotes the fair offer prescribes by the norm, and  $k_i$  is non-negative. Note, again, that  $k_1$  measures how much the proposer dislikes to deviate from what he takes to be the norm. To obey a norm, ‘sensitivity’ to the norm need not be high. Fear of retaliation may make a proposer with a low  $k$  behave according to what fairness dictates but, absent such risk, her disregard for the norm will lead her to be unfair.

Again, the responder should accept the offer if  $U_{2accept}(x) > U_{2reject} = 0$ , which implies the following *threshold for acceptance*:  $x > k_2 N / (1 + k_2)$ . Obviously the threshold is less than  $N$ : an offer more than what the norm prescribes is not necessary for the sake of acceptance.

For the proposer, the utility function is decreasing in  $x$  when  $x \geq N$ , hence a rational proposer will not offer more than  $N$ . Suppose  $x \leq N$ . If  $k_1 > 1$ , the utility function is increasing in  $x$ , which means that the best choice for the proposer is to offer  $N$ . If  $k_1 < 1$ , the utility function is decreasing in  $x$ , which implies that the best strategy for the proposer is to offer the least amount that would result in acceptance, i.e. (a little bit more than) the threshold  $k_2 N / (1 + k_2)$ . If  $k_1 = 1$ , it does not matter how much the Proposer offers provided the offer is between  $k_2 N / (1 + k_2)$  and  $N$ .

It should be clear at this point that  $k_1$  plays a very similar role as that of  $\beta_1$  in the Fehr-Schmidt model. In fact, if we take  $N$  to be  $M/2$  and  $k_1$  to be  $2\beta_1$ , the two models agree on what the proposer’s utility is. Similarly,  $k_2$  in this model is analogous to  $\alpha_2$  in the Fehr-Schmidt model. There is, however, an important difference between these formally analogous parameters. The  $\alpha$ ’s and  $\beta$ ’s in the Fehr-Schmidt model measure people’s degree of aversion towards inequality, which is a very different disposition than the one measured by the  $k$ ’s, i.e. people’s sensitivity to various norms. The latter will usually be a stable disposition, and behavioral changes may thus be caused by changes in focus or in expectations. A theory of norms can explain such changes, whereas a theory of inequity aversion does not. We will come back to this point later.

<sup>5</sup>Intuitively,  $N_2$  may be defined to proscribe rejection of fair (or hyper-fair) offers. The incorporation of this consideration, however, will not make a difference in the formal analysis.

<sup>6</sup>We assume there is no norm requiring that a responder ‘ought to’ reject an unfair offer.

It is also the case that the proposer’s belief about the responder’s type figures in her decision when  $k_1 < 1$ . The belief may be represented by a probability distribution over  $k_2$ . The proposer should choose an offer that maximizes the expected utility

$$EU(x) = P(k_2N/(1 + k_2) < x) \times (M - x - k_1(N - x)).$$

As will become clear, an advantage this model has over the Fehr-Schmidt model is that it can explain some variants of BUG more naturally. However, it shares a problem with the Fehr-Schmidt model in that they both entail that if the proposer offers a close-to-fair but not exactly fair amount, the only thing that prevented her from being too mean is the fear of rejection. This prediction, however, could be easily refuted by a parallel dictator game where rejection is not an option.

### 2.3 Reciprocity and fairness equilibrium: the Rabin model

The Fehr-Schmidt model does not consider reciprocity, a common phenomenon in human interaction, as we tend to be nice toward those who treated us well, and retaliate against those who slighted us. Matthew Rabin [1993] explicitly modeled reciprocity in the framework of psychological games [Geanakoplos *et al.*, 1989], introducing the important solution concept of *fairness equilibrium*.

The Rabin utility model is defined in a two-person game of complete information. The key idea is that a player’s utility is not determined solely by the actions taken, but also depends on the player’s beliefs (including second-order beliefs about first-order beliefs). Specifically, player  $i$  will evaluate her “kindness” to the other player,  $f_i$ , by the following scheme:

$$f_i(a_i, b_j) = \begin{cases} \frac{\pi_j(b_j, a_i) - \pi_j^c(b_j)}{\pi_j^h(b_j) - \pi_j^{min}(b_j)} & \text{if } \pi_j^h(b_j) - \pi_j^{min}(b_j) \neq 0 \\ 0 & \text{otherwise} \end{cases}, \quad i = 1, 2; j = 3 - i$$

where  $a_i$  is the strategy chosen by player  $i$ , and  $b_j$  is the strategy that player  $i$  believes is chosen by player  $j$ .  $\pi_j$  is  $j$ ’s material payoff function that depends on both players’ strategies.  $\pi_j^h(b_j)$  is the highest material payoff and  $\pi_j^{min}(b_j)$  the lowest payoff that player  $j$  can potentially get by playing  $b_j$ . In other words they denote, respectively, the highest and lowest payoffs player  $i$  can grant player  $j$  given that the latter is playing  $b_j$ . A key term here is  $\pi_j^c(b_j)$ , which is intended to represent the equitable material payoff player  $j$  deserves by playing  $b_j$ , and is defined by Rabin as:

$$\pi_j^c(b_j) = \frac{\pi_j^h(b_j) + \pi_j^l(b_j)}{2}$$

where  $\pi_j^l(b_j)$  is the worst payoff player  $j$  may incur given that player  $i$  does not play Pareto inefficient strategies. By definition,  $\pi_j^l(b_j) \geq \pi_j^{min}(b_j)$ .

Similarly, player  $i$  can estimate player  $j$ ’s kindness towards her, denoted by  $\tilde{f}_j$ :

$$\tilde{f}_j(b_j, c_i) = \begin{cases} \frac{\pi_i(c_i, b_j) - \pi_i^c(c_i)}{\pi_i^h(c_i) - \pi_i^{min}(c_i)} & \text{if } \pi_i^h(c_i) - \pi_i^{min}(c_i) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

where  $c_i$  is the object of a second-order belief — the strategy that player  $i$  believes player  $j$  believes to be chosen by  $i$ . The meaning of other terms should be obvious given the previous explanation. Finally, the utility function of player  $i$  is given by:

$$U_i(a_i, b_j, c_i) = \pi_i(a_i, b_j) + \tilde{f}_j(b_j, c_i) + \tilde{f}_j(b_j, c_i)f_i(a_i, b_j).$$

Care of reciprocity is reflected in the interaction term. Intuitively, it gives a player positive utility to be kind to (what she believes to be) kind players and tough to (what she believes to be) tough ones. An equilibrium of the game, called a *fairness equilibrium*, occurs when every belief turns out to be true, and each player’s strategy maximizes her utility, relative to the other player’s strategy and both players’ beliefs.

In order to apply the Rabin model to the BUG, we formulate the game in the following simultaneous fashion: player 1 (the proposer) chooses an amount as her offer and player 2 (the responder) chooses an amount as his threshold for acceptance. If the offer is lower than the threshold both get nothing, otherwise the money is divided accordingly. This way of looking at the game assumes that the responder does not adopt non-monotonic strategies in which some of the offers he would accept are lower than some of the offers he would reject. It also corresponds to the game of monopoly pricing that Rabin himself analyzed in his paper.

By definition, we have at equilibrium  $a_1 = b_1 = c_1$  and  $a_2 = b_2 = c_2$ , because all beliefs are true. Suppose  $a_1 = b_1 = c_1 = x$  and  $a_2 = b_2 = c_2 = y$ . That is, the proposer offers  $x$ , believes that the responder’s threshold is  $y$ , and believes that the responder believes she offers  $x$ ; the responder sets the threshold at  $y$ , believes that the proposer offers  $x$ , and believes that the proposer believes that the threshold is  $y$ . For what values of  $x$  and  $y$  is this a fairness equilibrium?

Let’s check. It is an equilibrium just in case  $a_1 = x$  maximizes  $U_1(a_1, b_2 = y, c_1 = x)$ , and  $a_2 = y$  maximizes  $U_2(a_2, b_1 = x, c_2 = y)$ . Consider  $U_1(a_1, b_2 = y, c_1 = x)$  first. Given that  $b_2 = y$ , offers less than  $y$  are Pareto inefficient options for player 1, because no player would be worse off and at least one player would be better off if player 1 chose an amount equal to or higher than  $y$ . It follows that  $\pi_2^l(b_2) = y$  and  $\pi_2^h(b_2) = M$ . Hence

$$\pi_2^c(b_2) = \frac{\pi_2^h(b_2) + \pi_2^l(b_2)}{2} = \frac{y + M}{2}$$

On the other hand, because all the Pareto efficient strategies given  $c_1$  result in  $(M-x, x)$  as the payoff,  $\pi_1^c(c_1) = M-x$ . Hence

$$f_1(a_1, b_2) = \frac{\pi_2(b_2, a_1) - \pi_2^c(b_2)}{\pi_2^h(b_2) - \pi_2^{min}(b_2)} = \begin{cases} -\frac{y+m}{2M} & \text{if } a_1 < y \\ \frac{2a_1 - (y+M)}{2M} & \text{if } a_1 \geq y \end{cases}$$

$$\tilde{f}_2(b_2, c_1) = \frac{\pi_1(c_1, b_2) - \pi_1^c(c_1)}{\pi_1^h(c_1) - \pi_1^{min}(c_1)} = \begin{cases} -1 & \text{if } x < y \\ 0 & \text{if } x \geq y \end{cases}$$

From  $\tilde{f}$  we see that the proposer never feels she is being treated “kindly”.<sup>7</sup> The utility function of player 1, as a function of  $a_1$ , is given by the following:

$$U_1(a_1, y, x) = \begin{cases} \frac{y-M}{2M} & \text{if } a_1 < y, x < y \\ 0 & \text{if } a_1 < y, x \geq y \\ M - a_1 - \frac{2a_1+M-y}{2M} & \text{if } a_1 \geq y, x < y \\ M - a_1 & \text{if } a_1 \geq y, x \geq y \end{cases}$$

It is not hard to see that when  $x < y$ ,  $U_1(x, y, x) < U_1(y, y, x)$  just in case  $M - y - y/M \geq 0$ . The latter holds unless  $y \geq M^2/(M + 1)$ , that is,  $y \approx M$ . Hence, unless the threshold is (almost) as high as the full amount,  $a_1 = x$  does not maximize  $U_1(a_1, b_2 = y, c_1 = x)$  if  $x < y$ , which means that  $(x, y)$  is not a fairness equilibrium.

On the other hand, when  $x \geq y$ ,  $U_1(a_1, b_2 = y, c_1 = x)$  is maximized at  $a_1 = y$ , and it is the unique maximum except when  $y = M$ . So, unless  $y = M$ , a necessary condition for  $(x, y)$  to be an equilibrium is that  $x = y$ .

Let’s now turn to player 2, the responder. The kindness functions are:

$$f_2(a_2, b_1) = \begin{cases} -1 & \text{if } x < a_2 \\ 0 & \text{if } x \geq a_2 \end{cases}$$

$$\tilde{f}_1(b_1, c_2) = \begin{cases} -\frac{y+M}{2M} & \text{if } x < y \\ \frac{2x-(y+M)}{2M} & \text{if } x \geq y \end{cases}$$

Based on this we can calculate player 2’s utility, as a function of  $a_2$ :

$$U_2(a_2, x, y) = \begin{cases} 0 & \text{if } x < a_2 \\ x - \frac{y+M}{2M} & \text{if } x \geq a_2, x < y \\ x + \frac{2x-(y+M)}{2M} & \text{if } x \geq a_2, x \geq y \end{cases}$$

As derived earlier, assuming  $y < M$ , it is necessary to have  $x = y$  for  $(a_1 = x, a_2 = y)$  to be an equilibrium. When  $x = y$ , we see that  $U_2(a_2, b_1 = x, c_2 = y)$  is maximized at  $a_2 = y$  just in case  $x + (x - M)/(2M) \geq 0$ , i.e., just in case  $x \geq M/(2M + 1)$ . Therefore, for any  $d \geq M/(2M + 1)$ , the following constitutes a fairness equilibrium:  $a_1 = b_1 = c_1 = a_2 = b_2 = c_2 = d$ .<sup>8</sup> If we take the unit to be one dollar, practically every offer is supported in some fairness equilibrium.

<sup>7</sup>This does not seem very plausible in the context of Ultimatum games, though it sounds all right when the game is phrased as monopoly pricing, as in Rabin’s analysis. However, we will see later that using a kindness function that avoids this consequence does not affect the equilibria.

<sup>8</sup>The condition  $d \geq M/(2M + 1)$  is equivalent to Rabin’s result about the highest price supported by fairness equilibrium in the context of monopoly pricing. The inequality  $M/(2M + 1) > 0$  corresponds to the fact that the highest price in Rabin’s result is less than the consumer’s valuation of the product, whereas the latter is the conventional monopoly price. Rabin regards this feature of the result a success of his model.

Of course, at all such equilibria the offer is accepted. When  $y = M$  and  $x < y$ , we see that  $U_2(a_2, b_1 = x, c_2 = y)$  is maximized at  $a_2 = y = M$  just in case  $x \leq 1$ . So there is no equilibrium at which the proposer offers more than 1 unit, but the responder rejects.

To summarize, there is a fairness equilibrium for practically any amount the proposer can offer at which the responder accepts the offer, and there is no equilibrium where an offer more than one unit is rejected. The first implication is undesirable for predictive purposes, and the second is problematic in view of the empirical data. One might suspect that these implications are due to the fact that Rabin’s definition of kindness is not suitable for the Ultimatum game. In particular, according to his definition, there are no “kind” responders from the proposer’s perspective.

There is a natural variant to Rabin’s definition of kindness that can avoid this counterintuitive consequence. Recall that all Pareto inefficient strategies are excluded in calculating the benchmark equitable material payoff  $\pi_j^c(b_j)$ . For extensive form games, however, one can naturally weaken the definition of Pareto efficiency: a strategy is Pareto efficient if it is not Pareto dominated in some subgames. Under this definition, given  $b_2 = y$ , offers lower than  $y$  are still Pareto efficient because they are Pareto efficient in some subgames (though they are not Pareto efficient along the equilibrium path), so  $\pi_2^l(b_2) = \pi_2^{min}(b_2) = 0$ . Hence  $\pi_2^c(b_2) = M/2$ . This subtle change affects the kindness functions. Suppose again  $a_1 = b_1 = c_1 = x$  and  $a_2 = b_2 = c_2 = y$ . Now the kindness functions for player 1— as functions of  $a_1$  — are the following:<sup>9</sup>

$$f_1(a_1, b_2) = \frac{\pi_2(b_2, a_1) - \pi_2^c(b_2)}{\pi_2^h(b_2) - \pi_2^{min}(b_2)} = \begin{cases} -\frac{1}{2} & \text{if } a_1 < y \\ \frac{2a_1 - M}{2M} & \text{if } a_1 \geq y \end{cases}$$

$$\tilde{f}_2(b_2, c_1) = \frac{\pi_1(c_1, b_2) - \pi_1^c(c_1)}{\pi_1^h(c_1) - \pi_1^{min}(c_1)} = \begin{cases} -\frac{1}{2} & \text{if } x < y \\ \frac{1}{2} & \text{if } x \geq y \end{cases}$$

(It is still the case that  $\pi_1^c(c_1) = M - x$ , even under the weaker condition for Pareto efficiency.) Observe that here  $\tilde{f}_2$  avoids the counterintuitive consequence that the proposer never feels being “kindly” treated by the responder. The proposer’s utility based on these kindness functions is given by:

$$U_1 a_1, y, x) = \begin{cases} -\frac{1}{4} & \text{if } a_1 < y, x < y \\ \frac{1}{4} & \text{if } a_1 < y, x \geq y \\ M - a_1 - \frac{M + 2a_1}{4M} & \text{if } a_1 \geq y, x < y \\ M - a_1 + \frac{M + 2a_1}{4M} & \text{if } a_1 \geq y, x \geq y \end{cases}$$

The implication of this utility function, however, is not different from our previous analysis. Again, if  $x < y$ , the utility cannot be maximized at  $a_1 = x$  unless  $y \approx M$ . On the other hand, when  $x \geq y$ , the utility is maximized at  $a_1 = x$  just in case

<sup>9</sup>Note that the kindness functions and the subsequent utility functions are still calculated according to Rabin’s scheme but are based on the new idea of Pareto efficiency of strategies.

$x = y$ . One can similarly recalculate the responder's utility function, and derive the same implications — that practically every offer is supported in some fairness equilibrium, and that there are no interesting equilibria in which the responder rejects an offer of more than one unit.

Therefore, the seemingly counterintuitive consequence of Rabin's kindness function is not responsible for the implications of Rabin's model. In our view, the problem lies in the fact that in Rabin's utility function, the relative importance of reciprocity/fairness versus material payoff is not well calibrated. The easiest way to fix this is to add a calibration parameter in the utility function, as used in many other utility models including the Fehr-Schmidt model and the Bicchieri model.

$$U_i(a_i, b_j, c_i) = \pi_i(a_i, b_j) + \alpha_i[\tilde{f}_j(b_j, c_i) + \tilde{f}_j(b_j, c_i)f_i(a_i, b_j)]$$

where  $\alpha_i$  measures the player's attitude towards the tradeoff between reciprocity/fairness and the material payoff. Using this utility function, we can run the same analysis as before, and derive the following:

1.  $a_1 = b_1 = c_1 = a_2 = b_2 = c_2 = d$  constitutes a fairness equilibrium just in case  $d \geq \alpha_2 M / (2M + \alpha_2)$ .
2. If  $\alpha_2 < M^2 / (M + \alpha_1)$ , then for every  $x \leq \alpha_2$  and  $y \geq M^2 / (M + \alpha_1)$ , the following constitutes a fairness equilibrium:  $a_1 = b_1 = c_1 = x, a_2 = b_2 = c_2 = y$ .

From (1), we see that given a high enough  $\alpha_2$ , many offers are not supported in any equilibrium in which the responder accepts the offer. In other words, if the responder cares enough about fairness, then the proposer's offer has to pass a significant threshold in order to reach an equilibrium at which the responder accepts. On the flip side, we see from (2) that there may be interesting equilibria in which the responder rejects a decent offer (more than one unit), depending on the type of the players.

### 3 VARIANTS TO BUG

So far we have only considered the basic Ultimatum game, which is certainly not the whole story. There have been a number of interesting variants of the game in the literature, to some of which we now apply the models to see if they can tell reasonable stories about the data.

#### 3.1 *Ultimatum Game with Asymmetric Information and Payoffs*

Kagel *et al.* [1996] designed an Ultimatum game in which the proposer is given a certain amount of chips. The chips are worth either more or less to the proposer than they are to the responder. Each player knows how much a chip is worth to her, but may or may not know that the chip is worth differently to the other. The

particularly interesting setting is where the chips have higher (three times more) values for the proposer, and only the proposer knows it. It turns out that in this case the offer is (very close to) half of the chips and the rejection rate is low. A popular reading of this result is that people merely prefer to *appear* fair, as a really fair person is supposed to offer about 75% of the chips.

To analyze this variant formally, we only need a small modification of our original setting. That is, if the responder accepts an offer of  $x$ , the proposer actually gets  $3(M - x)$  though, to the responder’s knowledge, she only gets  $M - x$ . In the Fehr-Schmidt model, the utility function of player 1 (the proposer), given the offer is accepted, is now the following:

$$U_{1\text{accept}}(x) = \begin{cases} (3 + 3\alpha_1)M - (3 + 4\alpha_1)x & \text{if } x \geq 3M/4 \\ (3 - 3\beta_1)M - (3 - 4\beta_1)x & \text{if } x < 3M/4 \end{cases}$$

The utility function of the responder upon acceptance does not change, as to the best of his knowledge, the situation is the same as in the simple Ultimatum game. Also, if the responder rejects the offer, both utilities are again zero. It follows that the responder’s threshold for acceptance remains the same: he accepts the offer if  $x > \alpha_2 M / (1 + 2\alpha_2)$ . For the proposer, if  $\beta_1 > 3/4$ , the best offer for her is  $3M/4$ , otherwise the best offer for her is the minimum amount above the threshold. An interesting implication is that even if someone offers  $M/2$  in the BUG, which indicates that  $\beta_1 > 1/2$ , she may not offer  $3M/4$  in this new condition. This prediction is consistent with the observation that almost no one offers 75% of the chips in the real game.

At this point, it seems the Fehr-Schmidt model does not entail a difference in behavior in this new game. But proposers in general do offer more in this new setting than they do in the BUG, which naturally leads to the lower rejection rate. Can the Fehr-Schmidt model explain this? One obvious way is to adjust  $\alpha_2$  so that the predicted threshold increases. But there seems to be no reason in this case for the responder to change his attitude towards inequality. Another explanation might be that under this new setting, the proposer believes that the responder’s distaste for inequality increases, for after all it is the proposer’s belief about  $\alpha_2$  that affects the offer. This move sounds as questionable as the last one, but it does point to a reasonable explanation. Suppose the proposer is uncertain about what kind of responder she is facing, and her belief about  $\alpha_2$  is represented by a non-degenerate probability distribution. She should then choose an offer that maximizes the expected utility, which in this case is given by the following:

$$EU(x) = P(\alpha_2 < x / (M - 2x)) \times ((3 - 3\beta_1)M - (3 - 4\beta_1)x)$$

This expected utility is slightly different from the one derived in the BUG in that it involves a bigger stake. As a result, it is likely to be maximized at a bigger  $x$  unless the distribution over  $\alpha_2$  is sufficiently odd. Thus the Fehr-Schmidt model can explain the phenomenon in a reasonable way.

If we apply the Bicchieri model to this new setting, again the utility function of player 1 (the proposer) does not change. The utility function of player 2 (the responder) does not change.

proposer) given acceptance is changed to

$$U_{1accept}(x) = 3(M - x) - k_1 \max(N' - x, 0)$$

We use  $N'$  here to indicate that the proposer's perception of the fair amount, or her interpretation of the norm, may have changed due to his awareness of the asymmetry. The model behaves pretty much in the same way as the Fehr-Schmidt one does. Specifically, the responder's threshold for acceptance is  $k_2 N' / (1 + k_2)$ . The proposer will/should offer  $N'$  only if  $k_1 > 3$ , so people who offer the "fair" amount in the BUG ( $k_1 > 1$ ) may not offer the "fair" amount under the new setting. That means even if  $N' = 3M/4$ , the observation that few people offer that amount does not contradict the norm-based model. The best offer for most people ( $k_3 < 3$ ) is the least amount that would be accepted. However, since the proposer is not sure about the responder's type, she will choose an offer to maximize her expected utility, and this in general leads to an increase of the offer given an increase of the stake. Although it is not particularly relevant to the analysis in this case, it is worth noting that  $N'$  is probably less than  $3M/4$  in the situation as thus framed. This point will become crucial in the games with obvious framing effects.

The Rabin model, as it stands, faces many difficulties. The primary trouble still centers on the kindness function. It is not hard to see that according to Rabin's definition of kindness, the function that measures the proposer's kindness to the responder does not change at all, while the function that measures the responder's kindness toward the proposer does change.<sup>10</sup> This does not sound plausible. Intuitively, other things being equal, the only thing that may change is the proposer's measure of her kindness to the responder. There is no reason to think that the responder's estimation of the other player's kindness toward him will change, as the responder does not have the relevant information. Strictly speaking, Rabin's original model cannot be applied to the situation where asymmetric information is present, because his framework assumes the payoffs are common knowledge. It is of course possible to adapt that framework to the new situation, but we will omit the formal details here.

It is, however, worth noting that if the kindness functions remain the same (as it is the case under our definition of kindness), the arguments available to Rabin to address the new situation are very similar to the ones available to the previous models. One move is to manipulate the  $\alpha$ 's, which is unreasonable as we already pointed out. Another move is to represent beliefs with more general probability distributions (than a point mass distribution), and to look for Bayesian equilibria. The latter strategy will inevitably further complicate the already complicated model, but it does seem to match the reality better.

---

<sup>10</sup>By our definition of kindness, both functions remain the same as in the simple setting.

### 3.2 *Ultimatum Game with Different Alternatives*

There is also a very simple twist to the Ultimatum game, which turns out to be quite interesting. Fehr *et al.* [2000] introduced a simple Ultimatum game where the proposer has only two choices: either offer 2 (and keep 8) or make an alternative offer that varies across treatments in a way that allows the experimenter to test the effects of reciprocity and inequity aversion on rejection rates. The alternative offers in four treatments are (5,5), (8,2), (2,8) and (10,0). When the (8,2) offer is compared to the (5,5) alternative, the rejection rate is 44.4%, and it is much higher than the rejection rates in each of the three alternative treatments. In fact, it turns out that the rejection rate depends a lot on what the alternative is. The rejection rate decreases to 27% if the alternative is (2,8), and further decreases to 9% if the alternative is (10,0).<sup>11</sup>

Is it hard for the Fehr-Schmidt model to explain this result? In their consequentialist model there does not seem to be any role for the available alternatives to play. As the foregoing analysis shows, the best reply for the responder is acceptance if  $x > \alpha_2 M / (1 + 2\alpha_2)$ . That is, different alternatives can affect the rejection rate only through their effects on  $\alpha_2$ . It is not entirely implausible to say that “what could have been otherwise” affects one’s attitude towards inequality. After all, one’s dispositions are shaped by all kinds of environmental or situational factors, to which the ‘path not taken’ seem to belong. Still it sounds quite odd that one’s sensitivity to fairness changes as alternatives vary.

The norm-based model, by contrast, seems better equipped to explain the data. For one thing, the model could explain the data by showing how different alternatives point the responder to different norms (or no norm at all). In particular, the way a situation is framed affects our expectations about others’ behavior and what they expect from us (Bicchieri 2006). For example, when the alternative for the proposer is to offer (5,5), players are naturally focused on an equal split. The proposer who could have chosen (5,5) but did not is sending a clear message about her disregard for fairness. If the expectation of a fair share is violated, the responder can be expected to resent it, and act to punish the mean proposer. In this case, a fairness norm applies, and indeed we observe 70% of proposers to choose (5,5) over (8,2). In the (8,2) vs. (2,8) case, 70% of proposers choose (8,2), since there is no norm (at least in our culture) saying that one has to sacrifice oneself for the sake of a stranger. When the choice is between (10,0) and (8,2), 100% of proposers choose (8,2), the least damaging (for the responder) outcome, and also what responders are believed to reasonably expect. Thus a natural explanation given by the Bicchieri model is that  $N$  changes (or even undefined) as the alternative varies.

A recent experiment in a similar spirit done by Dana, Weber and Kuang (2007) brings more light to this point. The basic setting is a Dictator game where the allocator has only two options. The game is played in two different situations. Under the “Known Condition” (KC), the payoffs are unambiguous, and the allo-

<sup>11</sup>Each player played four games, presented in random order, in the same role.

cator has to choose between (6, 1) and (5, 5), where the first number in the pair is the payoff for the allocator, and the second number is the other player's payoff. Under the "Unrevealed Condition" (UC), the allocator has to choose between (6, ?) and (5, ?), where the receiver's payoff is 1 with probability 0.5 and 5 with probability 0.5. Before the allocator makes a choice, however, she can choose to find out privately and at no cost what the receiver's payoff in fact is. It turns out that 74% of the subjects choose (5, 5) in KC, and 56% choose (6, ?) without revealing the actual payoff matrix in UC.

This result, as Dana *et al.* point out, stands strongly against the Fehr-Schmidt model. If we take the revealed preference as the actual preference, choosing (5, 5) in KC implies that  $\beta_1 > 0.2$ , while choosing (6, ?) without revealing in the UC condition implies that  $\beta_1 < 0.2$ . Hence, unless a reasonable story could be told about  $\beta_1$ , the model does not fit the data.<sup>12</sup> If a stable preference for fair outcomes is inconsistent with the above results, can a conditional preference for following a norm show greater consistency? Note that, if we were to assume that  $N$  is fixed in both conditions, a similar change of  $k$  would be required in the Bicchieri model in order to explain the data.<sup>13</sup>

However, the norm-based model can offer a more natural explanation of the data through an interpretation of  $N$ . In KC, subjects have only two, very clear choices. There is a 'fair' outcome (5,5) and there is an inequitable one (6,1). Choosing (6,1) entails a net loss for the receiver, and only a marginal gain for the allocator. Note that the choice framework *focuses* subjects on fairness though the usual Dictator game has no such obvious focus. Dana *et al.*'s example evokes a related situation (one that we frequently encounter) in which we may choose to give to the poor or otherwise disadvantaged: What is \$1 more to the allocator is \$4 more to the receiver, mimicking the multiplier effect that money has for a poor person. In this experiment, what is probably activated is a norm of beneficence, and subjects uniformly respond by choosing (5,5). Indeed, when receivers in Dana *et al.*'s experiment were asked what *they* would choose in the allocator's role, they unanimously chose the (5,5) split as the most appropriate.

A natural question to ask is whether we should hold the norm fixed, thus assuming a variation in people's sensitivity to the norm ( $k$ ), or if instead what is changing here is the perception of the norm itself. We want to argue that what changes from the first to the second experiment is the perception that a norm exists and applies to the present situation, as well as expectations about other people's behavior and what their expectations about one's own behavior might be [Bicchieri, 2006]. In Bicchieri's definition of what it takes for a norm to be

<sup>12</sup>In KC, choosing option B implies that  $U_1(5,5) > U_1(6,1)$ , or  $5 - \alpha_1(0) > 6 - \beta_1(5)$ . Hence,  $5 > 6 - 5\beta_1$  and therefore  $\beta_1 > 0.2$ . In UC, not revealing and choosing option A implies that  $U_1(6, (.5(5), .5(1))) > U_1(.5(5,5), .5(6,5))$ , since revealing will lead to one of the two 'nice' choices with equal probability. We thus get  $6 - .3(\beta_1) > 2.5 + .5(6 - \beta_1)$ , which implies that  $\beta_1 < 0.2$ .

<sup>13</sup>According to the Bicchieri model, if we keep  $N$  constant at 5, choosing option B in KC means that  $U_1(5,5) > U_1(6,1)$ . It follows that  $5 > 6 - 4k_1$ , i.e., that  $k_1 > 0.25$ . In UC, not revealing and choosing option A implies that  $U_1(6, (.5(5), .5(1))) > U_1(.5(5,5), .5(6,5))$ . It follows that  $6 - 2k_1 > 5.5$ , i.e., that  $k_1 < 0.25$ .

followed, a necessary condition is that a sufficient number of people expect others to follow it in the appropriate situations *and* believe they are expected to follow it by a sufficient number of other individuals. People will *prefer* to follow an existing norm *conditionally* upon entertaining such expectations. In KC, the situation is transparent, and so are the subjects' expectations. If a subject expects others to choose (5,5) and believes she is expected so to choose, she might prefer to follow the norm (provided her  $k$ , which measures her sensitivity to the norm, is large enough). In UC, on the contrary, there is uncertainty as to what the receiver might be getting. To pursue the analogy with charitable giving further, in UC there is uncertainty about the multiplier ("am I giving to a needy person or not?") and thus there is the opportunity for *norm evasion*: the player can avoid activating the norm by not discovering the actual payoff matrix. Though there is no cost to see the payoff matrix, people will opt to not see it in order to avoid having to adhere to a norm that could potentially be disadvantageous. So a person who chooses (5, 5) under KC may choose (6,?) under UC with the same degree of concern for norms. Choosing to reveal looks like what moral theorists call a *supererogatory* action. We are not morally obliged to perform such actions, but it is awfully nice if we do.

A very different situation would be one in which the allocator has a clear choice between (6,1) and (5,5), but she is told that the prospective receiver *does not even know* he is playing the game. In other words, the binary choice would focus the allocator, as in the KC condition, on a norm of beneficence, but she would also be cued about the absence of a crucial expectation. If the receiver does not expect the proposer to give anything, is there any reason to follow the norm? This is a good example of what has been extensively discussed in Bicchieri [2000; 2006]. A norm exists, the subject knows it and knows she is in a situation in which the norm applies, but her preference for following the norm is conditional on having certain empirical and normative expectations. In our example, the normative expectations are missing, because the receiver does not know that a Dictator game is being played, and his part in it. In this case, we would predict that a large majority of allocators will choose (6,1) with a clear conscience. This prediction is different from what a 'fairness preference' model would predict, but it is also at odds with theories of social norms as 'constraints' on action. One such theory is Rabin's [1995] model of moral constraints. Very briefly, Rabin assumes that agents maximize egoistic expected utility subject to constraints: Thus our allocator will seek to maximize her payoffs but experience disutility if her action is in violation of a social norm. However, if the probability of harming another is sufficiently low, a player may 'circumvent' the norm and act more selfishly. Because in Rabin's model the norm functions simply as a constraint, beliefs about others' expectations play no role in a player's decision to act. As the (6,1) choice does in fact 'harm' the recipient, Rabin's model should predict that the number of subject who choose (6,1) is the same as in the KC of Dana *et al.*'s experiment. In the Bicchieri model, however, the choices in the second experiment will be significantly different from the choices we have observed in Dana *et al.*'s KC condition.

### 3.3 *Ultimatum Game with Framing*

Framing effects, a topic of continuing interest to psychologists and social scientists, have also been investigated in the context of Ultimatum games. Hoffman *et al.* [1985], for instance, designed an Ultimatum game in which groups of twelve participants were ranked on a scale 1-12 either randomly or by superior performance in answering questions about current events. The top six were assigned to the role of “seller” and the rest to the role of “buyer”. They also ran studies with the standard Ultimatum game instructions, both with random assignments and assignment to the role of Proposer by contest. The exchange and contest manipulations elicited significantly lowered offers, but rejection rates were unchanged as compared to the standard Ultimatum game.<sup>14</sup>

Since, from a formal point of view, the situation is not different from that of the BUG, the previous analysis remains the same. Hence, within the Fehr-Schmidt model, one would have to argue that  $\alpha_2$  is decreased by the game framing. In other words, the role of a “buyer” or the knowledge that the proposer was a superior performer dampens the responder’s concern for fairness. This change does not sound intuitive, and demands some explanation. In addition, the proposer has to actually *expect* this change in order to lower her offer. It is equally, if not more difficult, to see why the framing can lead to different beliefs the proposer has about the responder.

In the Bicchieri model, the parameter  $N$  plays a vital role again. Although we need more studies about how and to what extent framing affects people’s expectations and perception of what norm is being followed, it is intuitively clear that framing, like the examples mentioned above, will change the players’ conception of what is fair. The ‘exchange’ framework is likely to elicit a market script where the seller is expected to try to get as much money as possible, whereas the entitlement context has the effect of focusing subjects away from equality in favor of an equity rule. In both cases, the perception of the situation has been changed, and with it the players’ expectations. An individual sensitivity to fairness may be unchanged, but what has changed is the salient division norm.

## 4 CONCLUSION

We have discussed how different utility functions try, with more or less success, to explain experimental result that clearly show that individuals take into account other parties’ utilities when making a choice. Material incentives are important, but they are just one of the items agents consider: the fairness of outcomes, the intentions of other parties, and the presence or absence of social norms are other important factors that play a role in decision-making. The three utility functions we have examined highlight different reasons why, in Ultimatum games, partici-

---

<sup>14</sup>Rejections remained low throughout, about 10%. All rejections were on offers of \$2 or \$3 in the exchange instructions, there was no rejection in the contest entitlement/divide \$10, and 5% rejection of the \$3 and \$4 offers in the random assignment/divide \$10.

pants tend to favor fair outcomes. However, the cross-situational inconsistencies that we observe in many variants of the Ultimatum game put to the test these different models. We believe a norm-based utility function can better explain the variance in behavior across experiments, but much more work needs to be done to design new experiments that directly test how much expectations (both normative and descriptive) matter, and when a norm is in fact present [Bicchieri and Chavez, 2010; Bicchieri and Xiao, 2009].

## BIBLIOGRAPHY

- [Bicchieri, 2000] C. Bicchieri. Words and Deeds: a Focus Theory of Norms. In *Rationality, Rules and Structure*. J. Nida-Rumelin and W. Spohn, eds., pp. 153-184. Dordrecht, Kluwer Academic Publishers, 2000.
- [Bicchieri, 2006] C. Bicchieri. *The Grammar of Society: The nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press, 2006.
- [Bicchieri and Xiao, 2009] C. Bicchieri and E. Xiao. Do the right thing: but only if others do so. *Journal of Behavioral Decision Making* **22**: 191-208, 2009.
- [Bicchieri and Chavez, 2010] C. Bicchieri and A. Chavez. Behaving as Expected: Public Information and Fairness Norms. *Journal of Behavioral Decision Making* **23**(2): 161-178, 2010.
- [Camerer, 2003] C. Camerer. *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton, NJ, Princeton University Press, 2003.
- [Dana et al., 2007] J. Dana, R. Weber, and J. X. Kuang. Exploiting Moral Wriggle Room: experiments demonstrating an illusory preference for fairness. *Economic Theory* **33**(1): 67-80, 2007.
- [Fehr et al., 1998] E. Fehr, E. Kirchler, A. Weichbold, and S. Gächter. When Social Norms Overpower Competition - Gift Exchange in Labor Markets, *Journal of Labor Economics* **16**, 324-351, 1998.
- [Fehr, 1999] E. Fehr and K. Schmidt. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics* **114**(3): 817-868, 1999.
- [Fehr et al., 2003] E. Fehr, A. Falk and U. Fischbacher. Testing Theories of Fairness - Intentions Matter. *Institute for Empirical Research in Economics, University of Zürich, Working Paper No. 63*, 2003
- [Geanakoplos et al., 1989] J. Geanakoplos, D. Pearce, et al. Psychological games and sequential rationality. *Games and Economic Behavior* **1**: 60-79, 1989.
- [Guth et al., 1982] W. Guth, R. Schmittberger, et al. An experimental analysis of ultimatum games. *Journal of Economic Behavior and Organization* **3**: 367-388, 1982.
- [Hoffman and Spitzer, 1985] E. Hoffman and M. Spitzer. Entitlements, rights, and fairness: An experimental examination of subjects' concept of distributive justice. *Journal of Legal Studies* **2**: 259-297, 1985.
- [Hoffman et al., 1994] E. Hoffman, K. A. McCabe, et al. Preferences, Property Rights, and Anonymity in Bargaining Games. *Games and Economic Behavior* **7**: 346-380, 1994.
- [Kagel et al., 1996] J. H. Kagel, C. Kim, et al. Fairness in ultimatum games with asymmetric information and asymmetric payoffs. *Games and Economic Behavior* **3**: 100-110, 1996.
- [Rabin, 1993] M. Rabin. Incorporating Fairness into Game Theory and Economics. *American Economic Review* **83**: 1281-1302, 1993.
- [Rabin, 1995] M. Rabin. Moral Preferences, Moral Constraints, and Self-Serving Biases, *Berkeley Department of Economics Working Paper No. 95-241*, August 1995.
- [Wicksteed, 1910] P. H. Wicksteed. *The Common Sense of Political Economy, Including a Study of the Human Basis of Economic Law*. London: Macmillan and Co, 1910.