

# Fair Procedures

## Evidence from Games Involving Lotteries

Gary E Bolton\*

Penn State University, USA

Jordi Brandts

Institut d'Anàlisi Econòmica (CSIC), Spain

Axel Ockenfels

Harvard University, USA & University of Magdeburg, Germany

July 2000

Procedural fairness plays a prominent role in the social discourse concerning the marketplace in particular, and social institutions in general. Random procedures are a simple case, and they have found application in several important social allocation decisions. We investigate random procedures in the laboratory. We find that an unbiased random procedure is an acceptable substitute for an unbiased allocation: similar patterns of acceptance and rejection result when either is inserted as a feasible proposal in a sequential battle-of-the-sexes. We also find that unbiasedness, known to be a crucial characteristic of allocation fairness, is important to procedural fairness: in the context of a random offer game, a biased outcome is more readily accepted when chosen by an unbiased random draw than by one that is biased. Procedural fairness is conceptually different than allocation fairness or attribution-based behavior, and none of the current models of fairness and reciprocity captures our results. Post hoc extension of one of these models (ERC) suggests that a deeper understanding of procedural fairness requires further investigation of competing fairness norms.

---

\* Correspondence. Bolton: Department of Management Science and Information Systems, 303 Beam, Penn State University, University Park, PA 16802, USA; geb3@psu.edu. Brandts: Institut d'Anàlisi Econòmica (CSIC), Campus UAB, 08193 Bellaterra, Barcelona, Spain; brandts@cc.uab.es. Ockenfels: Harvard University, Graduate School of Business Administration, Soldiers Field Road, Baker Library 188, Boston, MA 02163, USA; aockenfels@hbs.edu; and Faculty of Economics and Management, University of Magdeburg, Postfach 4120 D-39016 Magdeburg, Germany; axel.ockenfels@ww.uni-magdeburg.de.

We thank Sally Blount, and seminar participants in Bielefeld, Harvard and New York for helpful comments. We also thank David Rodriguez for help in running the experiments, and for skillful research assistance. Bolton gratefully acknowledges the support of the National Science Foundation, and Brandts that of the Spanish DGICYT (PB98-0465). All authors gratefully acknowledge the hospitality of the *Making Choices* project at the Center for Interdisciplinary Research, University of Bielefeld, during the summer of 2000.

Now the coat was without seam, woven from top throughout. They said therefore among themselves, Let us not rend it, but cast lots for it, whose it shall be. John 19:23-24

If a number of persons engage in a series of fair bets, the distribution of cash after the last bet is fair, or at least not unfair, whatever this distribution is. John Rawls (1971)

## 1. The observational importance of fair (random) procedures and two hypotheses

Economists have long been aware of the potential trade-off between market allocations that are efficient and those that are deemed fair. In the broader social discourse, however, market fairness is often discussed in a different light, in terms of whether the market constitutes a fair procedure. Searching the Internet on "fair procedures" turns up many and varied examples.<sup>1</sup> The same search illustrates that the social discourse on fair procedures goes beyond the marketplace, encompassing questions of justice (a fair trial) and democracy (a fair election).

But in what sense do people care about fair procedures? And by what criteria do people gauge whether a procedure is fair? A common observation is that those who complain about an unfair procedure tend to be the disaffected (e.g., dismissed employees, losing side in a trade dispute, least profitable firms). One might be tempted to conclude that "fairness" is little more than the rhetoric people use in pushing the outcome they favor. But this does not necessarily follow: for the disaffected may also *act* upon their dissatisfaction in ways that they would not if they deemed the process fair. They may, for instance, engage in acts of protest or obstruction.<sup>2</sup>

Work by experimental economists focused on *allocation fairness* provides indirect evidence in support of this last point. By allocation fairness we mean the preference people reveal for allocations they judge to be fair (we will deal with characterizing "fair" in a moment). The relevant finding is that, in situations such as ultimatum bargaining, people are often willing to incur a cost to resist allocations that are biased against them (Güth et al., 1982; Roth, 1995).

---

<sup>1</sup> A recent search turned up some 2000 hits. Among those relevant to the marketplace: A discussion on the Union of American Physicians and Dentists site on whether managed care networks have the right to terminate doctors absent a fair procedure. Remarks by the Chairman of the Federal Trade Commission on the need for industry self-regulating programs to abide by clear and fair procedures. A South African arbitration finding that a union employee was wrongly dismissed because the employer failed to follow fair procedures. A description posted by the Chilean government of the controls in place to ensure government procurement processes are fair. Remarks on the European Union server on the need for the World Trade Organization to develop fair procedures for dispute resolution. A Canadian university posting concerning what constitutes fair procedures for deciding tenure cases.

<sup>2</sup> There is some evidence in the psychology literature to suggest this. Greenberg (1990) reports a natural experiment in which employees in certain manufacturing plants had their pay temporarily cut by 15%. He found that employee theft rates rose by as much as 250% over those reported in a control condition with normal pay. In another treatment, the basis for the pay-cut was thoroughly explained to employees with the aim of reducing feelings of inequality, and the subsequent theft rate was reduced.

Recently, there has been theoretical progress in understanding this behavior, in the form of social utility models, of which it is common to distinguish two types. Relative payoff models (e.g., Bolton, 1991, Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000a) posit that people care about how their payoff compares to that of others, while intentionality models (e.g., Rabin, 1993) posit that people help or hurt those they believe intend to help or hurt them.

The experiments we report here focus on *procedural fairness*; so far as we know they are the first economics experiments to explicitly do so.<sup>3</sup> While the behavior we observe is in ways similar to that associated with allocation fairness, there are important differences. In fact, neither the relative payoff models nor the intentionality model explain the treatment effects we observe. These models nevertheless provide critical guidance to the investigation, and later we sketch an extension of one of the models to gain some insight into the workings of procedural fairness.

Our experiments involve random procedures. Several features recommend random procedures as a test case for ideas about procedural fairness. Random procedures are relatively simple, and are familiar to everyone who has ever made a decision by flipping a coin. Also, in a purely random procedure, the rules governing outcome selection are unambiguous. These rules can be precisely manipulated by simply altering the distribution from which the outcome is drawn. Finally, random procedures have a long and varied history of field use that, as we shall see, is closely associated with issues of fairness.

The hypotheses we study are motivated by observations made by scholars who have studied how random procedures work in the field. Elster (1989) provides a comprehensive discussion of the use of random procedures (lotteries) in social decision-making together with insightful examples. He notes that, “In most contexts, we want lotteries to be fair [...] in the sense of being truly random and unbiased” (p. 39). In modern times, random lotteries have found use in the awarding of oil drilling leases, admission to educational institutions, professional athletic drafts, tax auditing, allocation of public housing, allocation of scarce

---

<sup>3</sup> In contrast, there is a considerable body of work on procedural fairness in the experimental psychology literature, although we are not aware of any experimental psychology study that deals, as we do, with the case of random procedures. Most of the work in psychology focuses on determining what procedures or characteristics of procedures people judge to be fair, whereas we focus on how fairness effects people’s actions. Thibaut and Walker (1975) find that people’s judgements of third-party allocation and dispute resolution are influenced by the perceived fairness of the procedure, independent of how favorable the outcome. Tyler (1989) finds that the perceived neutrality, trust and deference of the decision-maker influences people’s judgement of procedural fairness. Lind and Tyler (1988) summarize much of the research.

medical resources such as kidney machines, as well as military drafts and jury selection (see Elster, 1989, chap. 2, for discussion and references).

Oberholzer-Gee, Bohnet and Frey (1997) present survey evidence of the social perception of the trade-off between efficiency and fairness in siting nuclear waste facilities. They find evidence that lotteries are acceptable to people if applied to a set of efficient options. To the list of field examples the authors add cellular telephone licenses, baggage inspection, as well as an amusing historical example: lotteries at the University of Basel to establish teaching assignments, resulting in the great mathematician Jakob Bernoulli teaching medicine.<sup>4</sup> The authors explain the social attraction to random procedures as follows (p. 89):

Random decision mechanisms are the embodiment of fair allocation procedures. None of the personal characteristics that typically interfere with decision processes in a completely unwarranted way enter procedures based on chance: Nepotism is out of the question. The rich and the powerful do not have any better chances than the poor and the humble if allocation relies on random decision processes. Provided the property rights initially assigned by a random mechanism can be freely transferred, there is not even a loss in allocative efficiency.

Kunreuther and Portney (1991) propose a two-stage process for siting noxious facilities, the first stage involving a lottery to address fairness concerns, and the second stage involving an auction to address efficiency concerns. The authors also describe an interesting case in which the Southeast Compact, a consortium of eight U.S. states, chose North Carolina by lottery to host a radioactive waste management facility. After some debate, the North Carolina legislature accepted the facility because they deemed the process fair.

A common rationale for the use of unbiased randomization in social-decision making is uncommonly well articulated in the judicial ruling handed down in *U.S. v. Holmes* in 1842. The case centered on a leaky and overcrowded lifeboat from which the crew chose to throw fourteen male passengers overboard. The judge found that the procedure used to choose the victims was flawed. Instead, he argued, the victims should have been chosen by a lottery, one in which the crew as well as the passengers participated. He said this would be “the fairest mode” because “In no other than this or some like way are those having equal rights put upon an equal footing, and in no other way is it possible to guard against partiality and oppression.”<sup>5</sup>

---

<sup>4</sup> Perhaps this is a good place to emphasize that our aim is not to justify the use of procedures that are fair but (perhaps grossly) inefficient. Instead, we seek to understand the *observational* importance of fair procedures, why these types of procedures play a prominent role in social dialogue and especially in practice.

<sup>5</sup> *U.S. v. William Holmes*, 1 Wallace Junior, 26 Fed. Cas. 30; also quoted by Elster (1989). The shipmen were found guilty of homicide.

Note two recurring themes in these examples: First, random procedures tend to be associated with goods or obligations exhibiting sharply restricted divisibility (e.g., athletic talent, nuclear waste facility).<sup>6</sup> Second, ‘fair’ randomization tends to be equated with ‘unbiased’ randomization; the judge’s words in the rowboat case provide eloquent justification. These observations motivate two hypotheses that we study in a controlled lab environment.

The first hypothesis we examine is that, when the divisibility of the good or obligation to be allocated is sharply restricted, a random procedure that is fair in the sense of being unbiased is an acceptable substitute for an allocation that is fair (unbiased). The important implication is that if a fair procedure is feasible, we should observe the same sort of pressure to select it as we do when a fair allocation is available. We investigate this hypothesis in the context of the battle-of-the-sexes played sequentially. This game has two efficient allocations, one biased in favor of the proposer, and one biased in favor of the responder. Recent studies of similar games find that the allocation favoring the proposer is rejected less often than when an equal division is also feasible.<sup>7</sup> (A common intuition is that this phenomenon discriminates between relative payoffs and intentionality as motivational triggers for rejection behavior, but this is in fact not clearly so.<sup>8</sup>) The data shows that when the proposer can select a lottery to decide between the biased allocations, the responder rejects the nonrandom selection of the biased allocation in much the same way as when an equal division is feasible. The lottery is accepted by nearly all.

Still, will a biased allocation chosen by random procedure be acceptable *ex post*, after it is clear who is on the short end of things? The North Carolina legislature’s decision suggests that this may be true – so long as the procedure is deemed fair. The second hypothesis then is that a biased allocation is more acceptable if chosen by an unbiased procedure than if chosen by one that is biased. We extend our experiment to study this hypothesis using the ultimatum game (it will be clear below that it is not practical to study this issue in the battle-of-the-sexes context). One of the important results about ultimatum games is that rejection rates drop when the

---

<sup>6</sup> This is not the only feature uniting many of these goods/obligations. Military drafts, health care, environmental degradation (nuclear waste facilities), and jury selection have all, at least at times, been associated with issues of justice or human rights. It is perhaps then not a coincidence that the issue of fairness looms particularly large for these allocation problems.

<sup>7</sup> Falk et al. (1999) and Bolton and Ockenfels (2000b); related findings are reported by Andreoni et al. (1999), Binmore et al. (1999), Brandts and Solà (forthcoming), and Yang et al. (1999).

<sup>8</sup> See the next section. In section 4, we argue that depending on the exact functional form of the *reference point* for fair outcomes or kind behavior, respectively, both classes of theories can in principle be brought in line with this observation.

proposal is made by a random draw (Blount, 1995). A common interpretation is that, with randomization, an unfair proposal can no longer be associated with the intentions of the proposer. In fact, while relative payoff models predict no difference, intentionality models suggest that rejections should fall. We reproduce the original result in the context of a mini-ultimatum game, using an unbiased randomization procedure. We go on to demonstrate that when the randomization is sharply biased in favor of the proposer, rejection rates rise up close to the level observed in the regular game.

In a post hoc analysis, we argue that a satisfactory explanation of the anomalous behavior in our experiment need draw on elements of both allocation and procedural fairness. We demonstrate how such an approach might work by sketching an extension of the ERC model (Bolton and Ockenfels, 2000a) that refines the relative payoff reference point. The exercise shows that a relative payoff model can, in principle, explain the phenomena. It also suggests the potential gains that might be had from a further investigation of competing fairness norms.

## 2. Study 1: Introducing a fair procedure into a sequential battle-of-the-sexes game

For the moment, we use the terms "fair" and "unbiased" interchangeably. Our aim for study 1 is to investigate the hypothesis that, when the divisibility of a good or obligation to be allocated is sharply restricted, a random procedure that is fair is an acceptable substitute for an allocation that is fair. We should then observe the same sort of pressure to select a fair procedure as we do when a fair allocation is available. The second study explores the extent to which unbiasedness is important to making a random procedure acceptable.

Figure 1. The three games of study 1

*Sequential Battle-of-the Sexes Game (BOS)*

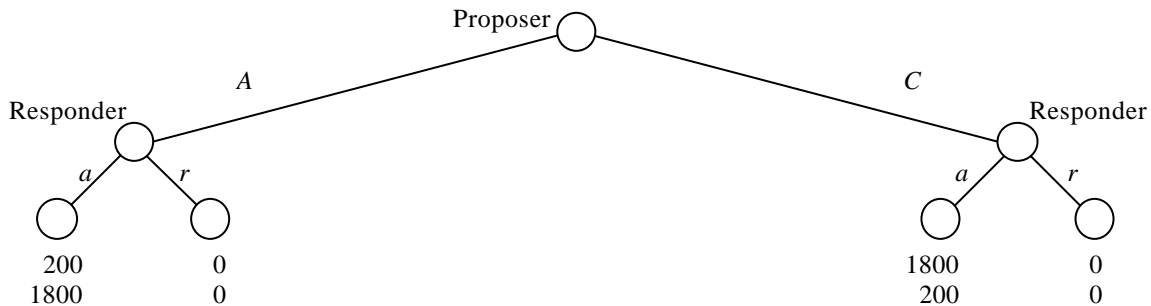
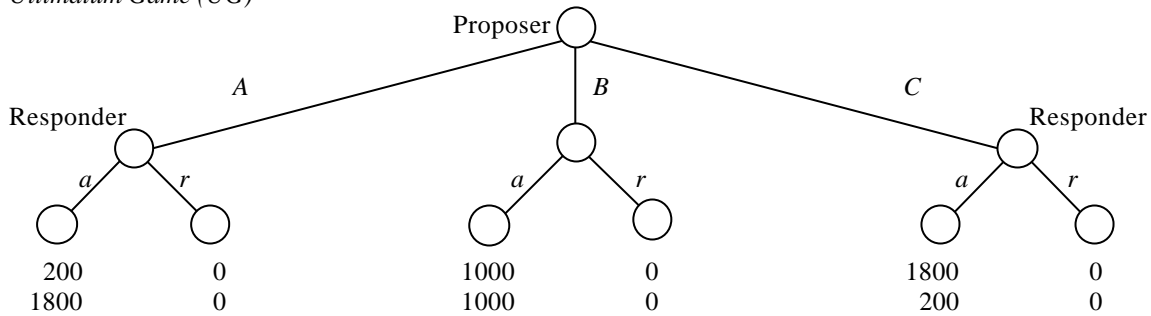
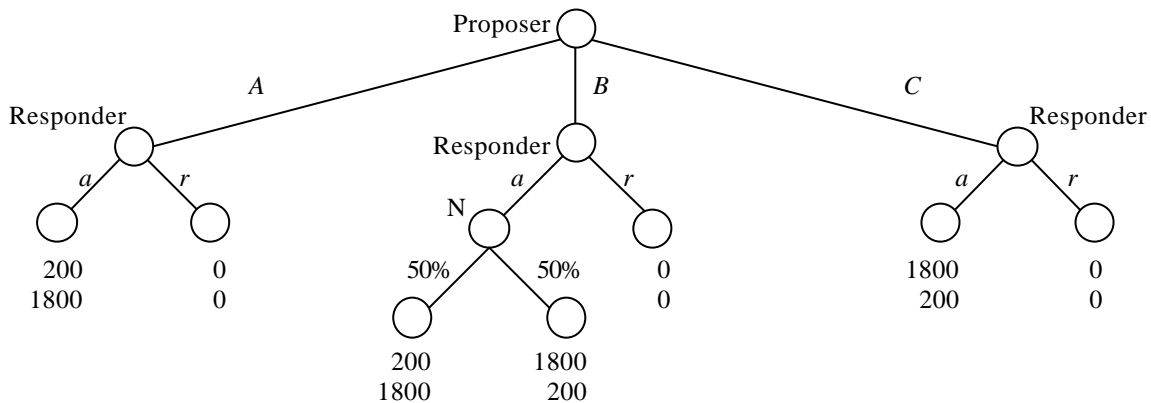


Figure 1 (continued). The three games of study 1

*Ultimatum Game (UG)*



*Sequential Battle-of-the-Sexes Game with Fair Procedure (BOSFP)*



The sequential version of the battle-of-the-sexes game (BOS) we study is illustrated in the first panel of Figure 1. The proposer chooses between a biased offer favoring self and a biased offer favoring the responder. The responder then accepts or rejects the proposer's offer. The experiment was run in Barcelona, and so the payoffs are in Spanish pesetas.<sup>9</sup>

The BOS captures a situation where the good or obligation cannot be allocated in an unbiased way. By adding the equal split offer of (1000,1000) we have a *mini-ultimatum game* (UG, panel 2 of Figure 1). Our test requires a third game, the *battle-of-the-sexes-with-fair-procedure* (BOSFP, panel 3 of Figure 1). This game permits the proposer the option of choosing an unbiased procedure – a lottery – to select the proposal. The standard perfect equilibrium for all games in Figure 1 has the proposer offering the (1800,200) split and the responder accepting.

The first fair procedure hypothesis implies that the rate of rejection of the (1800,200) proposal should be the same for both BOSFP and UG since the former exhibits a fair procedure

<sup>9</sup> At the time of the experiments, 2000 Spanish pesetas traded for about 12 U.S. dollars.

and the latter exhibits a fair allocation. Likewise, the fair procedure should be acceptable in the same manner that (we expect) the fair allocation to be acceptable.

Social utility models that focus on relative payoffs (Bolton, 1991, Fehr and Schmidt, 1999, and Bolton and Ockenfels, 2000a) predict no difference in the rate of rejection of (1800,200) offers across all three games. In these models, the relative payoff is a function of how the allocation compares to a relative reference point, and is independent of the set of feasible payoffs as well as the strategy space. The relative reference point for all of these models, when applied to 2-players games, is a 50-50 split.

As to intentionality, a common intuition is that, because the fair offer is not feasible in BOS, offering (1800,200) is less hostile than in UG, and should therefore be met with more kindness (less rejection). The intentionality model (Rabin, 1993) does not easily capture this intuition, however, because of the way kindness is measured. In essence, the kindness of the opponent's action is measured with respect to an absolute reference point that is a function of the most and least that a player believes that he could have received; relative payoffs or other interpersonal payoff comparisons do not play a role (note that they do, at least arguably, play a role in the common intuition). A straightforward application of this kindness measure suggests that an offer of 200 is evaluated as an equally hostile act in UG and BOS because in both games the same maximal offer of 1800 could have been chosen. The kindness of any response the responder could make to the (1800,200) offer, as well as the absolute consequences, is the same, implying similar rejection rates across the two games.<sup>10</sup>

It is conceivable that a somewhat modified reference point could, when inserted into the intentions model, capture the common intuition in a straightforward way. In section 4 we argue that this is true of the relative payoff models as well.

The fair procedures hypothesis makes no immediate prediction for the comparison of UG and BOSFP with the basic BOS although as we discuss later we think the hypothesis provides

---

<sup>10</sup> This illustration is somewhat simplified in order to make clear that the problem lies in the specific way the reference point is defined and not in the concept of intentionality *per se*. Technically, the model's measure of kindness also involves first and second order beliefs about what strategies will be taken. The argument in the text goes through so long as all players believe that all strategies are played with some (any) positive probability (this can be verified by inspecting the definitions of kindness given in Rabin, 1993). If some players *believe* that some strategies will be played with *zero* probability, then it is conceivable that the kindness measure may differ in the way necessary to deliver the result, although it is not immediately clear how this could be shown, or how robust such a result would be.



clues how to interpret what is actually observed. It is not entirely straightforward to say what the social utility models imply about the acceptance of the fair procedure, although if we evaluate the procedure at its expected value and assume risk neutrality then these models, like the fair procedure hypothesis, the models would seem to imply that the rate of acceptance should be the same as for a fair allocation.

### 2.1 Laboratory protocol

No subject played more than one of the three types of games. Procedures were the same for all games up to the necessary changes in the descriptions of the game actions and payoffs. The entire laboratory protocol appears in Appendix A.

The experiment employs a *no feedback design*, a method for eliciting the relevant responses from all subjects, free of historical contagion and so statistically independent across subjects. The trick is to have subjects play the game twice, once as proposer and once as responder, but not inform them of the outcome of the first game until the second is complete. To more closely simulate one-shot games, two additional precautions are taken. First, subjects are not matched with the same partner for the two games. Second, a coin flip after both games are completed determines just one game for actual payment. All of these procedures are public knowledge.<sup>11</sup>

All games were played in the normal form, also known as the strategy method; specifically, responders stated what they would do conditional on each of the proposer's choices prior to observing the proposer's actual choice (see Appendix A). This allows us to collect data on a player's complete strategy. It also adds comparability in that the benchmark experiment for study 2 (Blount, 1995) was played in this way.<sup>12,13</sup> There is some empirical evidence that the

---

<sup>11</sup> A no-feedback design does not investigate the effect of experience. Previous studies of games similar in nature to the ones we study here report no or small learning effects; e.g., Bolton and Zwick (1995), Kagel et al. (1996), Duffy and Feltovich (1999), Abbink et al. (2000), Bolton and Ockenfels (2000b), and Cooper et al. (2000). The latter authors find a small learning effect on the part of UG responders over an extended number of game repetitions. They affirm that, "no experiment to date has reliably observed learning on the part of [ultimatum game] responders." Our findings are based on rather large treatment effects, and a modest experience effect would not substantially alter our conclusions.

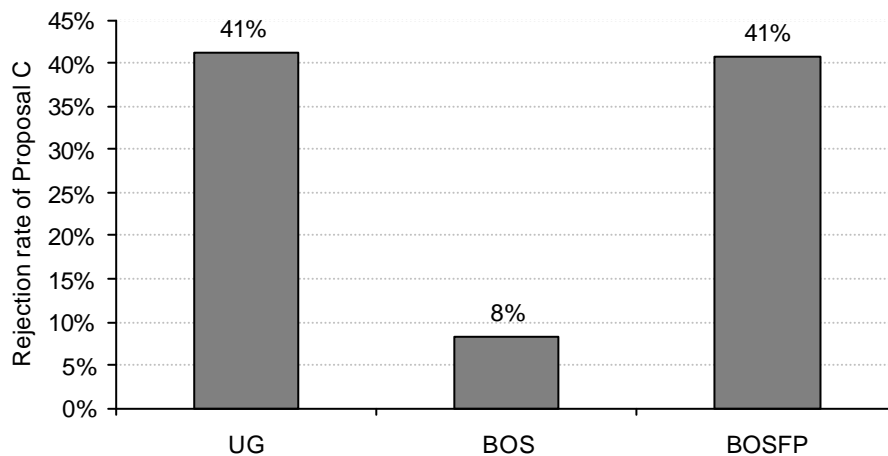
<sup>12</sup> Which is not to say that the designs are identical. Blount asked responders to name a minimum acceptance threshold over 21 proposals, whereas we have responders individually accept or reject each of three proposals. See section 3.

<sup>13</sup> One of the benchmark social utility models (Rabin, 1993) technically applies only to normal form games (but see Dufwenberg and Kirchsteiger, 1999, for an extensive-form version of Rabin's model).

form of play can make a difference (e.g., Schotter, Weigelt and Wilson, 1994). While there is no consensus on why this is so, a common conjecture is that the extensive form tends to evoke a hot (emotional) response, while the normal form tends to evoke a cool (studied) response. Applied here: normal form play might lessen overall rejection behavior, thereby dampening down treatment effects. The treatment effects behind the conclusions we draw, however, are all quite large (and in this sense the cool frame strengthens these results). And when a conclusion rests on the absence of a treatment effect, the absence is quite unambiguous.<sup>14</sup>

A total of 204 people participated in the experiments reported in the paper, with 102 participating in study 1. No subject participated in more than one session. Both BOS and BOSFP treatments had a sample size of 32, while the ultimatum treatment had a sample size of 38, differences in sample sizes reflecting differences in show-up rates. All subjects were students at the Universitat Autònoma de Barcelona. Data for each treatment was collected over two sessions. Each session ran approximately 40 minutes. Upon exiting, subjects were paid their earnings from the selected game plus a show-up fee of 500 pesetas.

Figure 2. Rejection rates of Proposal C in the ultimatum and battle-of-the-sexes games



---

<sup>14</sup> Brandts and Charness (2000) find little evidence for this conjecture in the context of two-player sequential prisoner's dilemma and chicken games, using a no-feedback design. Brosig, Weimann, and Yang (2000) find some supportive evidence in the context of sequential two-player bargaining experiments.

## 2.2 Results from study 1

The complete data set is broken out by aggregated proposer/aggregated responder and by individual subject in Appendix B. The main treatment effect is displayed in Figure 2. The main observations are

- When neither a fair procedure nor a fair outcome is feasible, an unfair offer is more acceptable: the rejection rates for Proposal *C* are higher in UG and in BOSFP than in BOS (one-sided  $\chi^2$ -test,  $p = 0.000$  for each comparison separately).
- A fair procedure may substitute for a fair outcome: (i) the rejection rate of Proposal *C* is virtually identical in BOSFP and UG (41.2% in UG and 40.6% in the BOSFP); (ii) the unbiased proposal (Proposal *B*) and the unbiased procedure are similarly acceptable. The unbiased proposal in UG was never rejected and the unbiased procedure in BOSFP was rejected just once.

It is also worth noting that Proposal *A* is rejected at a consistent rate across all three treatments (an average of 15% of the time), in spite of the fact that the proposal is biased in favor of the responder. One might conjecture that this represents mistakes on the part of confused or apathetic subjects. But if this were the case, we would expect to see similar rejection rates for the unbiased proposal; in fact the latter is never rejected. Likewise, we would expect to see similarly disperse proposer offers, but Proposal *A* was offered just once over the three treatments combined (a bit less than 1% of the time). Two alternative hypotheses suggest themselves. First, responders understood that Proposal *A* was very unlikely, and the rejections represent a form of cheap (fairman) talk. Second, rejection of Proposal *A* reflects strong inequality aversion, even in a favorable situation (as it is also expressed, yet in a weaker form, in dictator games). In other words, some subjects might prefer a fair but inefficient outcome to an allocation that implies favorable (or ‘exploitative’) inequality.<sup>15</sup>

---

<sup>15</sup> The relative payoff models of Fehr and Schmidt (1999) and of Bolton and Ockenfels (2000a) do not exclude the rejection of favorable inequality. There is, however, little empirical evidence on the subject, primarily for two reasons. First, proposers in ultimatum games typically do not offer more than half of the pie to the responders, so that favorable inequality aversion cannot be expressed. Second, if the game is played in normal (strategy) form representation, the experimenter often asks only for a ‘minimum acceptable offer’ so that only monotonic rejection strategies are observable (e.g., Blount, 1995).

Turning explicitly to proposer behavior, we find some differences across BOSFP and UG. BOSFP proposers are more likely to choose Proposal *C* than Proposal *B* in comparison to UG proposers (two-sided  $\chi^2$ -test,  $p = 0.051$ ). It is possible that BOSFP proposers (wrongly) anticipated lower rejection rates for Proposal *C* than UG proposers did, but it is also possible that risk plays a role. In particular, the fair allocation would be more attractive to risk-averse subjects than the lottery between the biased allocations. Consequently, proposers might be more readily willing to risk rejection in the fair procedures game. Also, all proposers in the BOS treatment select Proposal *C*. In the other two games, a majority chose Proposal *B* with all of the remainder save one choosing Proposal *C*. This is consistent with what is known as the I'm-no-saint effect (Bolton, Katok and Zwick, 1998): if a favored moderate outcome is removed, people tend to move to an outcome that favors self relative to the moderate outcome.

Before the experiments, we conjectured that individual behavior as a proposer and as a responder would be correlated. It is conceivable for instance, though it does not follow from any of the social utility models, that a proposer who chooses the fair offer is, as a responder, more likely to reject an unfair offer and vice versa. However, no statistically significant patterns relate proposer and responder role behavior (see the individual data in Appendix B).

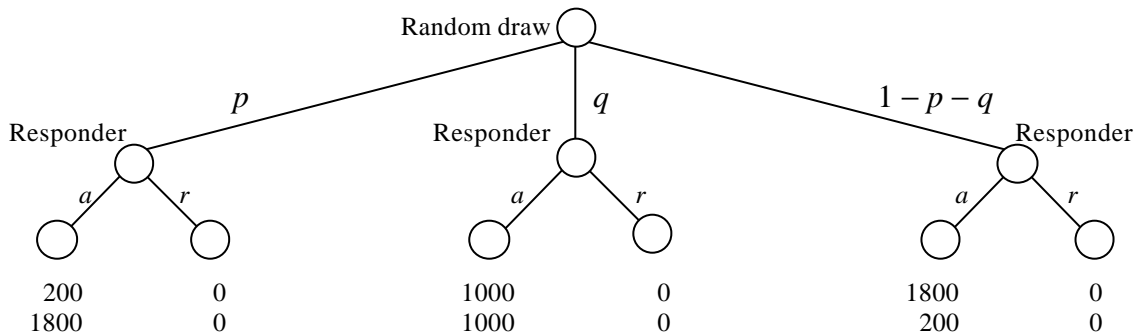
To summarize the main results, we find that unfair outcomes are more acceptable if neither procedural nor outcome fairness is feasible. We also find that a fair procedure can substitute for a fair proposal in the sense that both are similarly acceptable, and the availability of either make biased proposals similarly unacceptable.

### **3. Study 2: Biased and unbiased randomization in the mini-ultimatum game**

In study 2, we investigate whether the bias of the random procedure influences the *ex post* acceptability of the outcome of the procedure. The fair procedures hypothesis suggests that a biased allocation is more acceptable if chosen by an unbiased procedure than if chosen by a biased one. The *attribution hypothesis*, on the other hand, asserts that the important aspect of a random procedure is that a biased offer is no longer attributable to the proposer, and hence all allocations should be similarly acceptable independent of the bias of allocation or procedure.

We study these hypotheses in the context of the ultimatum game, for primarily two reasons. First, both the fair procedures and attribution hypothesis presuppose a substantial rate of rejection of biased offers in the non-randomization case. We have already seen that the battle-of-the-sexes, where no equal split nor random procedure is available, yields very low rejection rates for biased offers (8.3%), whereas the rejection rate for the ultimatum game is quite substantial (41%). The second reason is comparability. The well-known result that randomization can change rejection behavior is associated with the ultimatum game (Blount, 1995).<sup>16</sup> The original experiment compares an ultimatum game in which the proposer chooses the proposal in the usual way, with one in which the proposal is selected by an unbiased or nearly unbiased (at least by some measures) random draw.<sup>17</sup> Minimally acceptable offers in the latter type of game were roughly half what they were in the standard game.

Figure 3. Ultimatum games with random proposal



Treatments: (1) ASYM:  $(p, q, 1 - p - q) = (.01, .01, .98)$ , (2) SYM 98:  $(.01, .98, .01)$ , and (3) SYM34:  $(.33, .34, .33)$

To study 1, we add three additional treatments, all involving a version of the ultimatum game in which the proposal is chosen by a random throw of two ten-sided dice (Figure 3). The responder then either accepts or rejects the resulting draw. The payoff nodes are the same as for the UG. In the asymmetric treatment (ASYM), the probability density over proposals  $(A, B, C)$

<sup>16</sup> In a similar way, a random draw has been subsequently applied to the gift-exchange game; see Charness (1996).

<sup>17</sup> Blount (1995) exhibits the result with three random distributions. In her study 1, the distribution is perfectly symmetric about an offer of 50%, and so perfectly unbiased. In her study 2, the two distributions used are asymmetric, but the extent of the bias depends to a large extent on how it is measured. Measured by average, the bias is strong (average offers of 22% and 8%, respectively), but by other common measures, the bias is quite modest: both distributions have modal offers of 50%, one a median of 50% and the other a median of 45%.

is  $(p, q, 1-p-q) = (.01, .01, .98)$ , in the first symmetric treatment (SYM98), the probability density is  $(.01, .98, .01)$ , and in the second symmetric treatment (SYM34) we have  $(.33, .34, .33)$ . Two symmetric treatments permit a check for saliency effects. That is, if we found lower rejection rates for Proposal *C* in a single symmetric treatment, we might worry that there is more noise in symmetric responder rejection rates than in the asymmetric case since the probability of Proposal *C*, and therefore perhaps the associated payoff saliency, is lower in the symmetric case. By checking rejection rates across two symmetric treatments, we get a sense of whether this is a problem. It turns out that the rejection rates are identical, evidence that saliency is not an issue.

The fair procedures hypothesis implies the rejections of Proposal *C* (the biased offer) should be higher in ASYM than in the SYM treatments, with rates similar across the SYM treatments. The attribution hypothesis predicts that rejection rates should be uniformly low across all three new treatments.

As to the benchmark social utility models: The intentionality model implies that the responder rejects an offer to punish the bad intentions of the proposer. This model therefore implies the attribution hypothesis, and predicts that rejections of the unequal offer should be similar across all three randomization treatments but that this common rate should be lower than that in the UG. In contrast, the relative payoff models predict that rejection rates are determined by the division of the pie proposed, independent of how that division was arrived at. So rejection rates of Proposal *C* should be similar across all randomization treatments and the UG game.

### *3.1 Laboratory protocol*

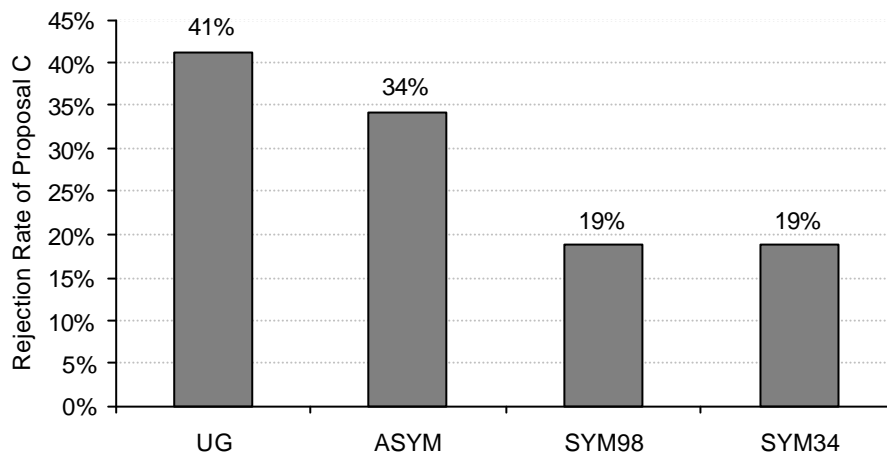
With the exception of the method for selecting a proposal, all other procedures were the same as for study 1 (see Appendix A). To select a proposal, the proposer threw two ten-sided dice. The experimenter supervised the throw and recorded the result. Responders did not observe the result, but did observe the monitor supervising.

The SYM98 and SYM34 treatments each had a sample size of 32, while the ASYM treatment had a sample size of 38 (differences in treatment sample sizes reflecting differences in show-up rates), for a total of 102 people. Subject pool demographics and payoff procedures were the same as for study 1.

### 3.2 Results from study 2

The complete data set appears in Appendix B. The main treatment effect is displayed in Figure 4:

Figure 4. Rejection rates of Proposal C in Ultimatum Games



Proposers make no selection in the new treatments, so all of our statistical observations have to do with responders:

- A fair procedure may substitute for an equal division: Rejection rates for Proposal C in treatments SYM34 and SYM98 are smaller than in the UG from study 1 (one-sided  $\chi^2$ -test,  $p = 0.024$ , for each comparison separately).
- The procedures that are unbiased have very similar effects: Rejection rates for Proposal C in treatments SYM34 and SYM98 do not differ (the rejection rate for both is 18.8%).
- There is no statistical difference in Proposal C rejection rates in the UG and in ASYM ( $p = 0.542$ , two-sided).
- An unfair procedure is not a substitute for a fair outcome: Rejection rates for Proposal C in treatment ASYM are higher than in treatments SYM34 and SYM98 ( $p = 0.074$  for each comparison separately; both our hypotheses and the data suggest that we can pool the data of the symmetric treatments, which then yields  $p = 0.040$ ).

The data for our SYM treatments agree with Blount's main observation that an unfair offer is more acceptable if chosen by a random draw than if chosen by the proposer. Turning to the other proposals (data in Appendix B): Proposal *B*, the unbiased proposal, is rejected just 4 times in total, and, similar to our first study (see the discussion in section 2.2), Proposal *A* is rejected 10 times (about 10%) in total. There is no statistically significant difference between number of rejections for Proposals *A* and *B*, respectively, across treatments regardless of whether the data of the symmetric treatments are pooled or not.

To summarize, our main observation is that, holding attribution constant, Proposal *C* is more acceptable if chosen by an unbiased random draw than if chosen by a biased random draw. The rejection rate after biased randomization is statistically indistinguishable from the rejection rate in the UG, while responder rejections for the unbiased random draws are substantially lower and very similar. The pattern of acceptance and rejection of Proposal *C* is in line with the procedural fairness hypothesis.

#### **4. A preliminary model of how outcome and procedural fairness interact**

Towards explaining the experiments, we might posit that what is really important to subjects is expected value. By this hypothesis, when responders in BOSFP play accept, what they are accepting is the expected value of the procedure. The important implication is that, if given the chance to respond to the biased allocation *after* the randomization, a higher proportion will reject the allocation than do the procedure. But from study 2, we see that the rate of rejection of Proposal *C* varies with the procedure used to select it (holding the set of potential allocations constant and controlling for the attribution hypothesis). This suggests that an explanation focused solely on allocation fairness will be unsatisfactory. We might then posit that the behavior we observe in the experiment, or at least that related to randomization, requires a wholly new explanation, separate from allocation fairness. But this approach also runs into problems: the equal division is virtually always acceptable, independent of how it is selected. It appears then that a satisfactory explanation of the experiment need incorporate elements of both procedural and allocation fairness.



In this section, we sketch an extension of the ERC model to address the findings of our experiments. Fehr and Schmidt's model could no doubt be extended in a similar way. The key insight is that changing the feasible outcome space, or determining the allocation by randomization, creates alternative fairness norms. We embed these norms into the model by refining the ERC reference point; once done, many of the observations from the present experiment fall in line. We caution from the start, that the model is extended post hoc, after the experiment was run and analyzed, and so the experiment cannot be considered a test of the model. Also, what we present here is a sketch. The intent is to demonstrate that a relative payoff model with a more refined reference point can, *in principle*, capture the phenomena in our experiment. We think this demonstration opens the way to a more careful consideration of how allocation and procedural fairness interact, and points to a role for competing fairness norms in understanding procedural fairness.

#### *4.1 The role of reference points in models of social utility*

The fact that the relative payoff models sometimes fail dramatically when the equal split is removed or chance moves are added has on occasion been interpreted as supportive evidence for intentionality. Contrary to this intuition, our study indicates that both classes of theories can fail. While one might interpret our results as evidence against the conceptual ideas underlying these models, we think that this conclusion is unnecessarily broad (nor does it offer a clear path to explaining the anomalies). Rather, we interpret our data as questioning the rudimentary reference points these models employ as measures of the fairness of an outcome, or alternatively, as measures for the kindness of an action. In particular, each of the game variations we study (removing equal outcomes and adding random procedures) establishes new, probably competing or intertwined, fairness norms that are not captured by the reference points used in the models.

As to removing the equal division, it seems plausible that whether an outcome is judged fair or whether an act is judged kind depends on how much fairness is feasible in the game. The reference points defined in Bolton and Ockenfels (2000a), Fehr and Schmidt (1999), and Rabin (1993) are invariant, or at least not very sensitive, to the removal of the equal division offer (see the discussion in section 2). A more sophisticated reference point for fair outcomes or kind

behavior might bring both classes of theories in line with the observation. Likewise, adding random procedures may introduce fairness norms specific to procedural fairness, something suggested by the Rawls quote at the beginning of the paper.

Our results imply that lower rejections in the random offer game are caused more by the fact that the procedure is fair than by differences in attribution – and this is one reason we focus our efforts on extending a relative payoff model. Another reason is that incorporating higher order beliefs into utilities, as required by intentionality models, substantially and, for our purposes, unnecessarily complicates the analysis. One might argue, doubting that a random procedure really controls *all* attribution-based explanations, that our biased random offer game triggers a relatively high rejection rate because responders punish the proposer for the unkind behavior of the *experimenter*.<sup>18</sup> Maybe this is true, but then again our experiment shows that whether the experimenter’s action is perceived as kind or not clearly depends on the *procedure* that she selects. Social utility models – whether they are fairness or attribution based or both – will not capture this sort of procedure-sensitive behavior so long as they only incorporate outcome-driven reference points.

#### 4.2 The generalized offer game

To put matters in a common framework, we first define a *generalized offer game*. There are two players: a proposer and a responder. The players jointly select a division of a pie normalized to size 1 from a finite and non-empty set,  $\mathbf{W}$ , of feasible allocation plans. If  $\mathbf{w} \in \mathbf{W}$  then either  $\mathbf{w}$  is an efficient allocation (i.e.,  $\mathbf{w}_1 + \mathbf{w}_2 = 1$ , subscripts denote which mover the share goes to), or  $\mathbf{w}$  is a lottery over efficient allocations (e.g., the lottery in BOSFP).

The game is played in two stages. In stage 1, the proposer selects a random procedure from a finite set  $\mathbf{L}$ ; each  $\mathbf{l} \in \mathbf{L}$  specifies probability weights over the allocation plans in  $\mathbf{W}$ . An offer of (1800,200) is treated as a lottery giving weight 1 to (.9,.1) and 0 to all other allocations

---

<sup>18</sup> Even though Blount did not mention this conjecture as an alternative explanation for her random offer game, she advances a similar line of argument to explain her third-party data. The latter shows that responders are rather insensitive to who makes the offer (also see section 5 of the present paper). This explanation contradicts the definition of the reference *group* given in the economic theories of intentionality (restricted to the players in the game who take some action), and on a deeper level, it appears to be at odds with the fundamental idea that motivated most of these theories, which is succinctly summarized in Rabin (1993, p. 1281): “People like to help those who are helping them, and to hurt those who are hurting them.” It is also unclear to us whether the ‘unkind experimenter’-hypothesis is testably different from the fair procedure hypothesis.

(payoffs stated as shares of the pie). The choice of  $\mathbf{I}$  leads to the realization of some  $\mathbf{w}$ , that is, uncertainty over  $\mathbf{W}$  specified by  $\mathbf{I}$  is resolved). In stage 2, the responder either accepts or rejects the resulting  $\mathbf{w}$ . If accepted,  $\mathbf{w}$  is realized and players are paid accordingly. If rejected, both players receive a payoff of 0.

The key to understanding this set-up is to note the careful distinction being made between the set of random procedures the proposer chooses from ( $\mathbf{L}$ ) and the set of allocation plans ( $\mathbf{W}$ ) that the responder could conceivably face to accept or reject. For example, for BOSFP the proposer chooses a procedure from a set with three elements: a lottery assigning weight 1 to (.9,.1), one assigning weight 1 to (.1,.9), and one assigning weight  $\frac{1}{2}$  to each possible allocation. The responder directly accepts or rejects the proposer's choice. So for this game,  $\mathbf{W} = \mathbf{L}$ . For SYM34, the proposer chooses from a singleton set (the fair lottery). After the procedure is realized, the responder accepts or rejects the selected allocation, meaning  $\mathbf{W} = \{(.9,.1), (.5,.5), (.1,.9)\}$ . The framework also accommodates the ultimatum game:  $\mathbf{L} = \mathbf{W} = \{(.9,.1), (.5,.5), (.1,.9)\}$ . For BOS,  $\mathbf{L} = \mathbf{W} = \{(.9,.1), (.1,.9)\}$ .

#### 4.3 An extension of ERC

Suppose that each player  $i$  acts to maximize the expected value of her utility function,

$$v_i = v_i(\mathbf{w}_i, \mathbf{s}_i).$$

The first argument,  $\mathbf{w}_i$ , specifies  $i$ 's absolute (monetary) payoff;  $v_i$  is non-decreasing in  $\mathbf{w}_i$ . The second argument, the relative payoff  $\mathbf{s}_i$ , reflects the social motivation. The functional form of  $\mathbf{s}_i$  has to be specified according to the appropriate social norm. For the moment, assume that  $v_i$  is concave down in  $\mathbf{s}_i$ , taking a unique maximum at  $\frac{1}{2}$  (i.e., the situation is perceived as 'fairest' if and only if  $\mathbf{s}_i = 1/2$ ):

$$v_{i1} \geq 0, v_{i2} = 0 \text{ if and only if } \mathbf{s}_i = 1/2, \text{ and } v_{i22} < 0.$$

If  $\mathbf{s}_i$  is measured as

$$\mathbf{s}_i = \mathbf{s}_i(\mathbf{w}_i^e) = \begin{cases} \mathbf{w}_i^e, & \text{if } \mathbf{w} \in \Omega \\ \frac{1}{2}, & \text{if } \mathbf{w} = (0,0) \end{cases}, \quad (1)$$

we have the ERC model proposed by Bolton and Ockenfels (2000a). The only difference is that equation (1) accounts for the fact that an outcome may be a lottery by evaluating the relative payoff at its expected value. The ERC model accounts for heterogeneity by permitting the trade-off between absolute and relative payoff to vary across players. Nevertheless, the equal division outcome plays a special role in that it maximizes everyone's relative payoff. This division can be thought of as a socially recognized fairness norm.<sup>19</sup>

Bolton and Ockenfels (2000a) show that formulation (1) of the relative payoff is consistent with regularities observed over a variety of games studied in the laboratory. In nearly all of these games, the equal split is a feasible outcome. It seems plausible that if the equal division, and therefore the simple fairness norm represented in (1), is not feasible, then some players might focus on a different norm (see the discussion in Bolton and Ockenfels, 2000a, p. 189; see Sen, 1997, for a general discussion of 'menu dependence' in preferences). An obvious choice would be the feasible outcome closest to equal division. Let  $m_d = \min_{\Omega} |\mathbf{w}_i^e - 1/2|$ .

For  $m_d < 1/2$ , define

$$\mathbf{s}'_i = \mathbf{s}'_i(\mathbf{w}_i^e, m_d) = \begin{cases} \mathbf{w}_i^e / (1 - 2m_d), & \text{if } \mathbf{w}_i^e \leq 1/2, \mathbf{w} \neq (0,0) \\ (\mathbf{w}_i^e - 2m_d) / (1 - 2m_d), & \text{if } \mathbf{w}_i^e > 1/2, \mathbf{w} \neq (0,0) \\ 1/2, & \text{if } \mathbf{w} = (0,0) \end{cases} \quad (2)$$

Since  $m_d = 1/2$  implies  $\mathbf{W} \subseteq \{(1,0), (0,1)\}$ , we define:  $\mathbf{s}'_i(\mathbf{w}^e, 1/2) \equiv 1/2$ . Equation (2) is a piece-wise linear rescaling of (1), so that  $\mathbf{s}'_i$  takes the value  $1/2$  at the feasible allocation closest to equal division; in effect, the allocation closest to equal division is taken as the social norm (when equal division is feasible, (1) and (2) are equivalent).<sup>20</sup> By formulation (2), in BOS both (1800,200)

---

<sup>19</sup> This does not mean that the model predicts that everyone will play fair – many people will want more than a fair share for its absolute payoff.

<sup>20</sup> In particular, this implies that for all  $\mathbf{w} \in \Omega$  and all  $m_d < 1/2$ :  $\mathbf{s}'_i(1, m_d) = 1$ ,  $\mathbf{s}'_i(0, m_d) = 0$ ,  $\mathbf{s}'_i(1/2 - m_d, m_d) = 1/2$ ,  $\mathbf{s}'_i(1/2 + m_d, m_d) = 1/2$ , and  $\mathbf{s}'_i(\mathbf{w}^e, 0) \equiv \mathbf{w}_i^e$ .

and (200,1800) are recognized as fair outcomes (both are the same minimum distance from equal division), whereas by formulation (1) neither is recognized as such.<sup>21</sup>

To add consideration of procedural fairness to (2), let  $\mathbf{I}_i^e$  be the expected value to  $i$  of the procedure the proposer chooses in stage 1. Then the relative payoff for  $i$  is either  $\mathbf{I}_i^e$  or  $\mathbf{S}_i'$  depending on whether the outcome or the procedure is less biased, i.e., which is closer to  $1/2$ :

$$\mathbf{S}_i'' = \mathbf{S}_i''(\mathbf{w}_i^e, m_d, \mathbf{I}_i^e) = \arg \min_{\mathbf{S}_i'(\mathbf{w}_i^e, m_d), \mathbf{I}_i^e} [|x - 1/2|] \quad (3)$$

By (3), the perception of fairness depends on player  $i$ 's share of the pie,  $w_i^e$ , on how much fairness is feasible,  $m_d$ , and on the (random) procedure that generates the outcome,  $\mathbf{I}_i^e$ . It can easily be checked that (3) is identical with the base ERC model formulation (1) if the equal split is feasible and random procedures are not an issue.

#### 4.4 Implications

The formulation in (3) implies that a biased outcome can be judged fair if there is no less biased outcome that is feasible *or* if the outcome was chosen by an unbiased procedure. For example, suppose  $i$  is the responder. Consider either of the SYM random offer games, and suppose the draw is  $\mathbf{w} = (.9, .1)$ . Then  $\mathbf{S}_i(.1) = \mathbf{S}_i'(.1, 0) = .1$ , but since  $\mathbf{I}_i^e = 1/2$  we have  $\mathbf{S}_i''(.1, 0, 1/2) = 1/2$ , and so  $v_i = v_i(.1, 1/2)$ . The value of rejecting is always  $v_i = v_i(0, 1/2) < v_i = v_i(.1, 1/2)$  implying the offer will be accepted independent of  $i$ . Now, suppose that the offer of (.9, .1) is generated by the ASYM random procedure. Again  $\mathbf{S}_i(.1) = \mathbf{S}_i'(.1, 0) = .1$ , but now  $\mathbf{I}_i^e = .11$  (computed using the ASYM probability weights) implying the offer may be rejected. On the other hand, by (3) a fair outcome (.5, .5) is never rejected, regardless of the procedure.<sup>22</sup>

Also, for BOSFP, offering (.9, .1) (with probability 1) implies  $\mathbf{S}_i(.1) = \mathbf{S}_i'(.1, 0) = .1$ ,  $\mathbf{I}_i^e = .1$ , and so  $v_i = v_i(.1, .1)$ , the same value as for the standard ultimatum game, and as in the

---

<sup>21</sup> Of course, there are plausible and more detailed alternative formulations of the relative payoff when the payoff space is restricted. For example, one might think that players have a more self-centered perception of fairness; e.g., they only care about how much fairness is feasible for *themselves* rather than how much fairness is feasible in the game. Since the amount of data for these sorts of games is at present modest, we choose to present the simplest formulation and leave the door open for refinements in the light of new evidence in other game variations.

<sup>22</sup> In line with this prediction, the fair outcome is rejected in less than 3% of all cases. Three out of the 5 rejections, however, occur in ASYM, suggesting that a small percentage of subjects are willing to reject the fair outcome when generated by an unfair process. The absolute numbers are too low, however, to conduct meaningful statistical treatment comparisons.

latter case, the proposal may be rejected. In contrast, accepting the proposal to randomize in BOSFP has a value of  $v_i = v_i(\mathbf{r}_i, \frac{1}{2})$  where  $\mathbf{r}_i$  denotes the absolute payoff certainty equivalent of the randomization to  $i$ . Given that the worst absolute payoff that might result from the lottery is .2, it is safe to assume that  $\mathbf{r}_i > 0$ , making  $v_i(\mathbf{r}_i, \frac{1}{2}) > v_i(0, \frac{1}{2})$ ; the proposal should be accepted.

If we suppose that all players gauge the fairness of an outcome by (3), we can explain the *comparative statics* of rejection – how rejection rates move or do not move across treatments in our experiment. Specifically, (3) implies that people reject the (1800,200) offer less often in BOS than in either UG or BOSFP where the equal split or a fair procedure, respectively, is feasible, and that the unfair offer is equally rejected in UG and BOSFP (see Figure 2). Furthermore, (3) implies that people more often reject the unfair offer in UG if selected by a biased procedure than if selected by an unbiased procedure. Selection of a biased outcome by either a proposer or a third party is treated as a biased procedure, and so UG rejection rates should be at least as high as for ASYM (see Figure 4).

However, the fact that there are *some* rejections in BOS and in the unbiased random offer games is not captured by reference point (3). A more realistic model would suppose that perceptions of fairness are divided across competing norms: some people judge fairness on simple outcome-based grounds (reference point measured by (1)), some take into account how much fairness is feasible in the whole game (2), and some additionally evaluate the procedure by which the outcome is determined (3). In line with our data, UG and ASYM would then have similarly high rejection rates of the (1800,200) offer (members subscribing to (1), (2) and (3) reject), SYM34 and SYM98 treatments lower but positive rates (members subscribing to only (1) and (2) reject), and BOS still lower rates (members subscribing to only (1) reject).<sup>23</sup>

The main insight here is that competing fairness norms, appropriately embedded in reference points, can explain much of the seemingly anomalous behavior we observe in our experiments on fair procedures. What we have sketched here is primarily responder behavior. A fuller, equilibrium description of the model requires a more explicit statement of how people

---

<sup>23</sup> Previous experimental evidence in similar environments strongly suggest that people differ in what they perceive as fair when there are multiple reasonable fairness norms; see e.g., Roth and Malouf's (1979) early studies of Nash-bargaining, Babcock et al.'s (1995) study of the self-serving bias in fairness judgements, and Ockenfels and Weimann's (1999) study of systematic differences in fairness notions across cultures.

choose between competing fairness norms, or at least need be precise about how people are distributed across norms. Incorporating this sort of heterogeneity into the model is not, in principle, different from how the individual heterogeneity already in the model is handled (see Bolton and Ockenfels, 2000a).

## 5. Conclusions

The ‘equity’ of procedures is a central theme in the development of many modern social institutions: equal opportunity in the marketplace, equal rights under the law, one-man-one-vote. In this sense, our results on procedural fairness, centered as they are on unbiasedness, speak to a general phenomenon.<sup>24</sup> Our data indicates that people respond to procedural fairness issues much as the hypotheses we began with suggest: The opportunity for a fair procedure has much the same effect on the acceptability of a given allocation as does the opportunity to have a fair outcome. Results produced by an unbiased procedure tend to be more acceptable than those produced by a biased procedure.

These findings begin to explain the pressure to implement social procedures that people broadly deem fair. Simply put, some people are willing to pay a cost – and subsequently impose a cost on others – to resist procedures that they deem to be biased against them.

We argued in the last section that the results of our experiment require an explanation that meshes procedural fairness norms with allocation fairness norms, and demonstrated how this approach might work by sketching an extension of a relative payoff model so that it reflects competing fairness norms. The extension tracks the data well. The approach nevertheless raises some concerns. As Binmore et al. (1999) state, in speaking about social utility explanations in general, “[T]he results will be useful only if some portability of the preferences can be recovered, in the form of some systematic view of the relationship between the specification of the game and preferences.” We agree. At the same time, our model sketch demonstrates that the search for a systematic manipulation of preferences might fruitfully be narrowed to a search for a

---

<sup>24</sup> While the data from this experiment is consistent with the notion that people have a preference for egalitarianism (a strong version of relative payoff preferences), it is also consistent with a more self-centered concept, that of a preference for equity for self. Likewise, the appeal of, for example, the principle of “one-man-one-vote,” can be construed as an appeal to egalitarianism, but can also be construed as an appeal to each individual’s self-centered concern for equity. See the related discussion in the next footnote.

systematic manipulation of the reference point. Different situations might systematically evoke different reference points (fairness norms), so that a practical taxonomy could be developed by identifying natural classes of games to which they apply.

While refining the reference points in social utility models appears to be a promising research path, even our most complete model does not capture the heterogeneity of individuals' perceptions of what is fair in our games, nor do we have reason to believe that it is general enough to capture different fairness norms that might emerge in other, possibly more complex games. Of course, one could take the view that what is needed is an altogether different conceptual approach. But if it is true that different games systematically trigger different perceptions of what is fair, it is hard to imagine *any* general model that will explain the facts without addressing the heterogeneity of fairness norms. And if there is no such thing as a universal fairness or reciprocity norm to guide social behavior independent of the game, any model with a simple statement of those norms is bound to be incomplete. This does not mean, of course, that such models are useless. It has been shown that we can go a surprisingly long way with very simple models of fairness in some important classes of games. For more challenging games, as shown in section 4, systematic variations of fairness norms can be incorporated into these models in a way that allows subsequent empirical tests and theoretical refinements.<sup>25</sup>

One of the intriguing issues that our experiment raises is 'credibility.' Think of a seaman claiming to have randomly selected the passenger to be thrown overboard, or a judge assuring a defendant that due process was followed, or a politician claiming that his leadership is the will of the people, or a proposer in a BOS game claiming to have arrived at his offer of (1800,200) by mentally flipping a coin. Without *visible* proof that these decisions were arrived at by a fair procedure, it is hard to imagine the disaffected – doomed passengers, convicted defendants,

---

<sup>25</sup> It is known that competing fairness norms arise in games with more than two players. The fairness models by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000a), for instance, disagree on the reference point for these games. While Fehr and Schmidt (1999) assume that players have egalitarian preferences, Bolton and Ockenfels (2000a) choose a simpler 'self-centered fairness' formulation that presumes that players only care about their own relative status. The evidence in games with more than two players is mixed (see the discussion and references in Bolton and Ockenfels, forthcoming), supporting the conclusion that there is no reason to suppose that there is a single simple fairness reference point. Both theories suggest that in many competitive market situations, fairness does not influence behavior, independent of the relative reference point used.



alienated citizens, or losing responders – accepting it. Only observable and verifiable unbiased procedures would seem *credibly* fair.

Blount (1995) also studied a third party treatment in which the offer is determined by a third person with no self-interest in the outcome of the game. She could not find a substantial difference in rejection rates between the standard ultimatum treatment and the third party treatment: “Perhaps, the most interesting finding of Studies 1 and 2, is the fact that the third party condition elicited a retaliatory response similar to that observed in the interested party condition.” (p.139; only in one experiment, in her Study 3 involving a special frame, did she find a difference). It is plausible that an unknown third party making the decision by unknown criteria lacks the credibility of observable randomization. A firm conclusion requires further investigation.

## References

- Abbink, K., Bolton, G.E, Sadrieh, K. and Tang, F. (2000), "Adaptive Learning Versus Punishment in Ultimatum Bargaining," working paper, Penn State University.
- Andreoni, J., Brown, P.M., Vesterlund, L. (1999), "What makes an Allocation Fair? Some Experimental Evidence," working paper, University of Wisconsin, Madison.
- Babcock, L., Loewenstein, G., Issacharoff, S., and Camerer, C. (1995), "Biased Judgements of Fairness in Bargaining," *American Economic Review*, 85(5), 1337-1343.
- Binmore, K., McCarthy, J., Ponti, G., Samuelson, L. and Shaked, A. (1999), "A Backward Induction Experiment," SSRI working paper 9934R, University of Wisconsin, Madison.
- Blount, S. (1995), "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences," *Organizational Behavior and Human Decision Processes*, 63, 131-144.
- Bolton, G.E (1991), "A Comparative Model of Bargaining: Theory and Evidence," *American Economic Review*, 81, 1096-1136.
- Bolton, G.E, Katok, E. and Zwick, R. (1998), "Dictator Game Giving: Rules of Fairness versus Acts of Kindness," *International Journal of Game Theory*, 1998, 27, 269-299.
- Bolton, G.E and Ockenfels, A. (2000a), "ERC: A Theory of Equity, Reciprocity and Competition," *American Economic Review*, 90(1), 166-193.
- Bolton, G.E and Ockenfels, A. (2000b), "Evidence on the Importance of Reference Points in Social Utility Models," working paper, Penn State U., <http://silmaril.smeal.psu.edu:80/lems/>.
- Bolton, G.E and Ockenfels, A. (forthcoming), "Self-centered Fairness in Games with More than Two Players," in: Charlie Plott and Vernon Smith (eds.), *Handbook of Experimental Economics Results*, <http://silmaril.smeal.psu.edu:80/lems/>.
- Bolton, G.E and Zwick, R. (1995), "Anonymity versus Punishment in Ultimatum Bargaining," *Games and Economic Behavior*, 10, 95-121.
- Brandts, J. and Charness, G. (2000), "Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games," *Experimental Economics*, 2, 227-238.
- Brandts, J. and Solà, C. (forthcoming), "Reference Points and Negative Reciprocity in Simple Sequential Games," *Games and Economic Behavior*.
- Brosig, J., Weimann, J., and Yang, C. (2000), "Emotions and Reasons in a Hot versus Cold Experiment," working paper, University of Magdeburg.

- Charness, G. (1996), "Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation," working paper, University of California, Berkeley.
- Cooper, D.J., Feltovich, N., Roth, A.E. and Zwick, R. (2000), "Relative versus Absolute Speed of Adjustment in Strategic Environments: Responder Behavior in Ultimatum Games," working paper, Case Western University.
- Duffy, J., and Feltovich, N. (1999), "Does Observation of Others Affect Learning in Strategic Environments? An Experimental Study," *International Journal of Game Theory*, 28(1), 131-152.
- Dufwenberg, M., and Kirchsteiger, G. (1999), "A Theory of Sequential Reciprocity," working paper, University of Stockholm.
- Elster, J. (1989), *Solomonic Judgements*, Cambridge: University Press.
- Falk, A., Fehr, E., and Fischbacher, U. (1999), "On the Nature of Fair Behavior," working paper, University of Zürich.
- Fehr, E., and Schmidt, K. (1999), "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 817-868.
- Greenberg, J. (1990), "Employee Theft as a Reaction to Underpayment Inequity: The Hidden Cost of Pay Cuts," *Journal of Applied Psychology*, 75, 561-568.
- Güth, W., Schmittberger, R., and Schwarze, B. (1982), "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, 3, 367-388.
- Kagel, J., Kim, C., and Moser, D. (1996), "Fairness in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs," *Games and Economic Behavior*, 13, 100-110.
- Kunreuther, H., and Portney, P. (1991), "Wheel of Fortune: A Lottery/Auction Mechanism for Siting of Noxious Facilities," *Journal of Energy Engineering*, 117, 125-132.
- Lind, E.A., and Tyler, T.R. (1988), *The Social Psychology of Procedural Justice*, New York: Plenum Press.
- Oberholzer-Gee, F., Bohnet, I., and Frey, B.S. (1997), "Fairness and Competence in Democratic Decisions," *Public Choice*, 91, 89-105.
- Ockenfels, A., and Weimann, J. (1999), "Types and Patterns – An Experimental East-West-German Comparison of Cooperation and Solidarity," *Journal of Public Economics*, 71(2), 275-287.
- Rabin, M. (1993), "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83, 1281-1302.

- Rawls, J. (1971), *A Theory of Justice*, Cambridge, MA: Belknap Harvard Press, (quote at top of paper from p. 75 of revised edition 1999, paperback).
- Roth, A.E. (1995), "Bargaining Experiments," In J. Kagel and A.E. Roth (eds.), *Handbook of Experimental Economics*, Princeton: Princeton University Press.
- Roth, A.E., and Malouf, W.K. (1979), "Game-theoretic Models and the Role of Information in Bargaining," *Psychological Review*, 574-594.
- Schotter, A., Weigelt, K., and Wilson, C. (1994), "A Laboratory Investigation of Multiperson Rationality and Presentation Effects," *Games and Economic Behavior*, 6, 445-468.
- Sen, A. (1997), "Maximization and the Act of Choice," *Econometrica*, 65, 745-779.
- Thibault, J. and Walker, L. (1975), *Procedural Justice: A Psychological Analysis*, Hillsdale, N.J.: Erlbaum.
- Tyler, T.R. (1989), "The Psychology of Procedural Justice: A Test of the Group-Value Model," *Journal of Personality and Social Psychology*, 57, 830-838.
- Yang, C., Mitropoulos, A., and Weimann, J. (1999), "An Experiment on Bargaining Power in Simple Sequential Games," working paper, University of Magdeburg.

## Appendix A. Laboratory Protocol

This section contains a description of all procedures, as well as verbal and written instructions given to subjects for the BOS treatment. Other treatments were run with minimal changes. (The randomization for the study 2 games was accomplished using two ten-sided dice.) The monitor read all verbal instructions directly from the protocol. The only monitor-subject communication not included in the protocol are answers to individual subject questions (answers were given in private). The positive and negative reciprocal action treatments differed only as indicated on written instructions and task forms. The protocol has been translated from Spanish. The Spanish text is available from the authors upon request.

*Seating.* Upon entering the room each participant is randomly directed to a seat. Half the seats have red folders and half have blue. Red folder seats are on the right side of the room, and blue folder seats on the left side. May I have your attention please. We are ready to begin. Thank you for coming. With the exception of the folder, please remove all materials from your desk. Open your folder and take out the sheet marked 'Instructions'. At this time please read the instructions. *Participants read silently.*

### *Instructions*

*General Information.* The purpose of this session is to study how people make decisions in a particular situation. If you have any questions, feel free to raise your hand and a monitor will assist you. From now until the end of the session, unauthorized communication of any nature with other participants is prohibited.

During the session you will play a game that gives you an opportunity to make money. Upon completion of the session the amount you make will be paid to you in cash. Payments are confidential: no other participant will be told the amount of money you make.

*Description of the game.* The game involves two players, *A* and *B*. There are three possible proposals for how *A* and *B* might divide 2000 pesetas

Proposal 1: *A* receives 200 and *B* receives 1800.

Proposal 2: *A* receives 1800 and *B* receives 200.

Proposal 3: The throw of a coin selects between proposal 1 and proposal 2.

The game has two steps.

Step 1: *A* selects the proposal of his choosing.

Step 2: *B*, who does not know the proposal selected in step 1, states whether he or she will accept or reject each of the three possible proposals.

After steps 1 and 2 are complete, the monitors will match the results of the two steps. If the selected proposal was accepted, the money is divided among *A* and *B* according to the selected proposal. If the selected proposal was rejected, both *A* and *B* receive zero for the game. In the case that proposal 3 was made and accepted, we will throw a coin to select either proposal 1 or proposal 2.

*Development of the session.* You will participate in two games, one played after the other. Both games are identical to the description in the previous paragraph. For each game, you will be paired with a different person. No one, however, will know the identity of the persons they are paired with, nor will these identities be revealed after the session is complete.

In game 1, those with red folders will have the role of *A*, and those with blue folders will have the role of *B*. For game 2, blue folders are *A*, and red folders are *B*. The results for game 1 will not be revealed prior to completion of game 2.

*Payment.* You will be paid 500 pesetas for your participation plus your earnings for one of the two games. The game that will be paid will be selected by a coin flip after both games have been completed. You should therefore play each game as if it is the game to be paid. Once you are paid, you may leave.

After a few minutes a monitor starts reading the instructions. I will now read the instructions out loud. During the reading the steps of the game are shown on an overhead projector that can be seen by all participants. Are there any questions?

*Decision making procedure.* We are ready to begin game 1. Those with red folders will receive the form for role A, and those with blue folders will receive the form for role B. In step 1, one of the three proposals for dividing 2000 pesetas is selected. A monitor passes out forms for game 1 to the subjects in the player A role. Players A please review the form with me. The form is projected on an overhead projector and read to the subjects.

### **Game 1: Form for player A**

You have the role of A. The person you are paired with has the role of B.

Step 1 of the game. Please select one of the three proposals by circling the proposal of your choice:

Proposal 1: A receives 200 and B receives 1800.

Proposal 2: A receives 1800 and B receives 200.

Proposal 3: The throw of a coin selects between proposal 1 and proposal 2.

When finished, please turn this sheet over and wait quietly. *The written form ends here.*

*Subjects are given time to make their decisions. Then a monitor collects the forms.* Step 1 is now complete. A monitor hands out forms for game 1 to the subjects in the player B role. In step 2, players B indicate whether they will accept or reject each possible proposal. Players B please review the form with me. *The form is projected on an overhead projector and read to the subjects.*

### **Game 1: Form for player B**

You have the role of B. The person you are paired with has the role of A.

Step 2 of the game. Indicate whether you will accept or reject each of the three possible proposals by circling either accept or reject for each one.

Proposal 1: accept    reject

Proposal 2: accept    reject

Proposal 3: accept    reject

When finished, please turn this sheet over and wait quietly. *The written form ends here.*

*Subjects are given time to make their decisions. Then a monitor collects the forms.* We are now ready to begin game 2. Remember, you will be paired with a different person than you were in game 1. Game 2 is then conducted using the same procedure as for game 1. Once complete, a coin is flipped to determine the game for payment. For the cases where proposal 3 has been selected by A and accepted by B a coin is flipped to determine the outcome. Then subjects are paid and leave the room.

## Appendix B: Complete Data Set

### *Aggregated Data*

<i>Ultimatum Games</i>	Nobs		Proposal A (200, 1800)	Proposal B (1000, 1000)	Proposal C (1800, 200)
UG	34	# offers	1	23	10
		# rejections	6	0	14 (41.2%)
SYM34	32	# rejections	1	1	6 (18.8%)
SYM98	32	# rejections	3	0	6 (18.8%)
ASYM	38	# rejections	6	3	13 (34.2%)
<hr/>					
<i>Battle of the Sexes Games</i>	Nobs		Proposal A (200, 1800)	Proposal B $Pr\{A\} =$ $Pr\{C\} = 50\%$	Proposal C (1800, 200)
BOS	36	# offers	0	not applicable	36
		# rejections	4		3 (8.3%)
BOSFP	32	# offers	0	14	18
		# rejections	5	1	13 (40.6%)

*Individual data*

<i>Ultimatum Game (no randomization)</i>		Number of proposers who proposed ...		
What is rejected?	Proposal A	Proposal B	Proposal C	
nothing	1	11	8	
<i>C</i>	0	6	2	
<i>A and C</i>	0	6	0	
Sum	1	23	10	

<i>Ultimatum Game with random proposal</i>		Number of proposers who proposed ...		
What is rejected?	ASYM	SYM98	SYM34	
nothing	21	26	27	
<i>C</i>	11	3	4	
<i>A and C</i>	2	3	1	
<i>B and C</i>	0	0	1	
<i>A and B</i>	3	0	0	
<i>A</i>	1	0	0	
Sum	38	32	32	

<i>BOS</i>		Number of proposers who proposed ...		
What is rejected?	Proposal A	Proposal B	Proposal C	
nothing	0	not	29	
<i>A</i>	0	applicable	4	
<i>C</i>	0		3	
Sum	0		36	

<i>BOSFP</i>		Number of proposers who proposed ...		
What is rejected?	Proposal A	Proposal B	Proposal C	
nothing	0	13	3	
<i>C</i>	0	4	6	
<i>A and C</i>	0	0	3	
<i>B</i>	0	0	1	
<i>A</i>	0	1	1	
Sum	0	18	14	