The Economic Journal, **115** (*October*), 1054–1076. © Royal Economic Society 2005. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

FAIR PROCEDURES: EVIDENCE FROM GAMES INVOLVING LOTTERIES*

Gary E Bolton, Jordi Brandts and Axel Ockenfels

Procedures are the area where fairness arguably has its largest influence on modern societies. The experiments we report provide an initial characterisation of that influence and suggest new interpretations for some well-known results. We find that procedural fairness is conceptually distinct from allocation fairness, although the evidence also indicates that the two are linked in important ways. *Post hoc* extension of one of the current models of fairness illustrates this link and implies that a deeper understanding of procedural fairness will require investigation of competing fairness norms.

> 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)

Economists have long been aware of the potential trade-off between market allocations that are efficient and those deemed fair. In the broader social discourse, 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.[1] The same search illustrates the discourse beyond the marketplace, encompassing questions of justice (a fair trial) and democracy (a fair election). In modern societies, procedures deemed fair are typically those that create a 'level playing field', a place where the participants have equal opportunity even if the resulting allocation is not equal.

We report here on a series of experiments designed to provide an initial characterisation of the influence procedural fairness has on choice behaviour. So far as

* We thank Sally Blount, two anonymous referees, and seminar participants in Ann Arbor, Bielefeld, Carnegie Mellon, Harvard, Magdeburg, New York, Tilburg and Zürich for very 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, Brandts that of the Spanish Ministry of Education and Culture (SEC-2002-01352) and the Barcelona Economics Programme of CREA, and Ockenfels that of the German DFG. All authors gratefully acknowledge the hospitality of the Making Choices project at the Centre for Interdisciplinary Research, University of Bielefeld, during the summer of 2000.

[1] A recent Google search turned up some 2.4 hits (July 2nd, 2004). Among those relevant to the marketplace were:

A discussion on the Union of American Physicians and Dentists site of whether managed care networks have the right to terminate doctors' contracts in the absence of a fair procedure.

Remarks by the Chairman of the Federal Trade Commission on the need for industry self-regulating programmes to abide by fair procedures.

A South African arbitration finding that a union employee was wrongly dismissed because the employer failed to follow fair procedures.

A description 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 Organisation to develop fair procedures for dispute resolution.

A Canadian university posting concerning what constitutes fair procedures for deciding tenure cases.

we know, these are the first economics experiments of this kind.[2] In contrast, a considerable amount of work has been directed at allocation fairness (Kagel and Roth, 1995). One of the critical findings in this line of research is that fair outcomes are often enforced by resistance to unfair outcomes. The classic example is the two-person ultimatum bargaining game. Ultimatum offers of small portions of the pie are commonly rejected, even though both bargainers would benefit materially from acceptance. This behaviour drives average offers towards equal division outcomes.[3]

Our study revolves around random procedures of a sort familiar to anyone who has ever resolved a difference with a coin flip. We insert random procedures into the ultimatum game and the closely associated battle-of-the-sexes as a mechanism for choosing between outcomes. We study the influence fair procedures have on the pattern of acceptance and resistance to various outcomes. In study 1, we ask players to commit to the results of the procedure, *prior* to executing the procedure, when there is uncertainty about the allocation to be chosen. In study 2, we ask for *ex post* commitment to the individual possible results. In both cases, we find strong effects on the pattern of acceptance and resistance. One way to summarise the significance of our results is to note that social utility models,[4] which collectively do a pretty good job of explaining allocation fairness, nevertheless do not anticipate our findings and, in some cases, offer what would otherwise be regarded as compelling intuition against them. (Later we will argue that a satisfactory explanation of our findings requires a unified model of both allocation and procedural fairness that might be derived by extending existing social utility theory.) *Unbiased* random procedures capture the equal opportunity or 'level playing field' element that appears critical to many of the procedures that modern societies deem fair. In fact, random lotteries have a long history of field use and tend to be used to distribute goods or obligations for which equal division among all participants is not practicable, and when there is no clear alternative distribution criterion. For example, in California, in case of ties in elections of town mayors, the law calls for the winner to be decided by chance if there are no special town rules set up to handle ties. In December 2002 two mayoral candidates tied with 546 votes in the town of Waterford. A deck of cards was shuffled, and the candidates cut the cards to determine who should be mayor.[5] In modern times, random lotteries have been used in the allocation of public housing and scarce medical resources, the awarding of oil drilling leases, admission to educational institutions, professional

[2] There is a large body of work on procedural fairness in the psychology literature. Most of this work is focused on determining what procedures people judge to be fair, whereas we focus on how fairness affects choice (although, of course, there is a perceptions aspect to choice). Thibault and Walker (1975) find that people's judgments of third-party allocation and dispute resolution are influenced by the perceived fairness of the procedure, independent of how favourable the outcome. Tyler (1989) finds that the perceived neutrality, trust and deference of the decision maker influences people's judgment of procedural fairness. Lind and Tyler (1988) and Tyler and Lind (2000) summarise much of the research

[3] Forsythe *et al.* (1994) provides one of the clearest empirical demonstrations of this principle.

[4] E.g., Bolton (1991), Rabin (1993), Fehr and Schmidt (1999), Dufwenberg and Kirchsteiger (forthcoming), Bolton and Ockenfels (2000), Charness and Rabin (2002).

[5] CBS News commented 'No costly run-off elections, no appeals to the courts, just a cut of the cards. What could be fairer?' http://www.cbsnews.com/stories/2002/12/11/opinion/garver/main532631.shtml.

athletic drafts, tax auditing, as well as military drafts and jury selection (Elster, 1989).[6] Oberholzer-Gee *et al.* (1997) present survey evidence that lotteries are an acceptable procedure for siting nuclear waste facilities if applied to a set of efficient options. They explain the social attraction to random procedures this way:

> 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.

Random procedures also play to the strength of the lab (control) in that they can be precisely manipulated, and the bias or lack of bias can be cleanly conveyed to all subjects.

The particular hypotheses we examine are motivated by anecdotal evidence suggestive of the influence that the fairness or unfairness of a procedure can have on choice. One anecdote, concerning the *ex ante* acceptability of a procedure, suggests a hypothesis that runs counter to the intuition offered by social utility models. The second anecdote, concerning the *ex post* acceptability of an outcome, suggests a new explanation for a well-known ultimatum game result, one that is often interpreted independent of the fairness of the random procedure involved.

The common rationale for the *ex ante* desirability of unbiased randomisation is uncommonly well articulated in the judicial ruling handed down in US *v. Holmes* in 1842. The case centred on a leaky and overcrowded lifeboat from which the crew chose to throw fourteen male passengers overboard. The shipmen were ultimately found guilty of homicide, although not for throwing people overboard – this was accepted as necessary to save some lives – rather because the procedure they used to choose the victims, one that exempted the crew, was judged *ex ante* unacceptable. In ruling, the judge argued that the victims should have been chosen by a lottery in which both crew and 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.'[7]

The hypothesis we investigate in our first study is that, when a fair (unbiased) procedure is feasible and a fair allocation is not, selection of a fair procedure is acceptable whereas there is resistance to the imposition of biased outcomes, much as when a fair (unbiased) allocation is available. We investigate the hypothesis in the context of the sequential battle-of-the-sexes, a game with two efficient allocations, each biased in favour of one of the players. One reason for wondering about the

[6] Oberholzer-Gee *et al.* (1997) add an amusing historical example: lotteries at the University of Basel to determine teaching assignments, resulting in the great mathematician Jakob Bernoulli teaching medicine. We note that our aim is not to justify normatively the use of procedures that are fair but (sometimes grossly) inefficient. Instead, we seek to understand the observational importance of these procedures in social practice.

[7] US *v. William Holmes*, 1 Wallace Junior, 26 Fed. Cas. 30. While the court ruled *ex post* it took an *ex ante* perspective.

general validity of this hypothesis, is that the use of the procedure in no way mitigates the outcome inequality – someone is ultimately on the same short end of things with or without the procedure; in fact, essentially for this reason, social utility models predict that the introduction of an unbiased random procedure to our game will have little positive *ex ante* consequence. Nevertheless, our data are largely in line with the fair procedure hypothesis. (Later, we will argue that this evidence provides insight into how people choose between competing procedural and outcome norms.)

Still, will a biased allocation chosen by random procedure be acceptable after it is clear who is on the short end of things? The Southeast Compact, a consortium of eight US states, chose North Carolina by lottery to host a radioactive waste management facility. After some debate, the North Carolina legislature accepted the facility, *ex post*, because they deemed the process fair (Kunreuther and Portney, 1991). The second hypothesis then is that a biased allocation is more acceptable when chosen by an unbiased procedure than when chosen by a biased one. We extend our experiment to study this hypothesis in the context of a random offer ultimatum game (it will be clear below that it is not practical to use the battle-of-the-sexes).[8] One of the important results about ultimatum games is that rejection rates drop when the proposal is made by random draw (Blount, 1995). A common interpretation is that, with randomisation, an unfair proposal can no longer be associated with the intentions of the proposer. We reproduce the original result in the context of a mini-ultimatum game, using an unbiased randomisation procedure. We go on to demonstrate that when the randomisation is sharply biased in favour of the proposer, rejection rates rise close to the level observed in the regular game – something that all social utility models get wrong. At the same time we find that an unbiased outcome will be accepted independent of whether the procedure that led to it was biased or not.

If we had done only study 1, we might have concluded that the pattern of resistance to unfair outcomes is the same in the presence of either a fair procedure or a fair outcome. In contrast, if we had only done study 2, we might have concluded that the pattern of resistance is fundamentally different since biased outcomes that are unacceptable in the allocation case are acceptable when chosen by a fair procedure. But when put together, the two studies suggest that allocation and procedural fairness interact in a subtle but important way. In a *post hoc* analysis, we sketch an extension of a social utility model that is roughly consistent with what we see in our data. The critical approach we advance is that refining the reference point may prove adequate to account for fair procedures.

1. Study 1: *Ex ante* Acceptance of a Fair Procedure in a Sequential Battle-of-the-sexes Game

Our first study involves the sequential version of the battle-of-the-sexes game (BOS) and the mini-ultimatum game (UG), both displayed in Figure 1 (payoffs for

[8] A similar hypothesis has recently been tested in a clever field experiment by Frey and Stutzer (2001). They note the lack of empirical work in economics on what they call ‘procedural utility.’ Their study indicates that people gain procedural utility from participating in the political decision-making process itself (measured by individuals’ reported subjective well-being), irrespective of outcome. Two other useful studies in the area are Anand (2001) and Frey *et al.* (2002).

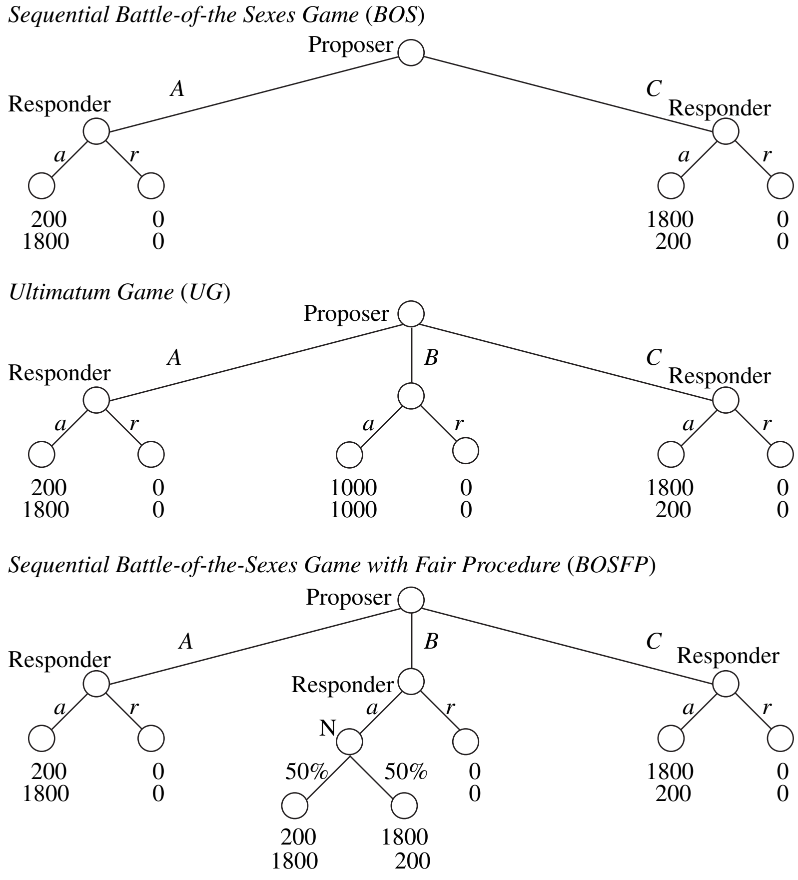


Fig. 1. *The Three Games of Study 1*

the game are in Spanish pesetas.[9]) In the BOS, the proposer chooses between a biased offer favouring self and a biased offer favouring the responder. In the UG, the proposer can additionally make an equal division offer of (1000, 1000). In both games, upon receiving the offer, the responder either accepts or rejects. (We will return to discuss the third game in the figure in a moment.)

One of the important findings from past studies of games like these is that the allocation favouring the proposer is rejected less often when an equal division is absent, as in BOS, than when it is feasible, as in UG.[10] We will find the same. One interpretation of this result is that 'what-is-fair' depends on how much outcome fairness is possible; so (1800, 200) is more acceptable in BOS than in UG because in the former game, a more equal division is not possible. This sort of 'menu

[9] The experiment was run in Barcelona during the year 2000, prior to the introduction of the euro. At the time of the experiments, 2000 Spanish pesetas traded for about 12 US dollars or 12 euros.

[10] Falk *et al.* (2001), Güth *et al.* (2001) and Bolton and Ockenfels (forthcoming *a*); related findings are reported by Andreoni *et al.* (2002), Binmore *et al.* (2002), Brandts and Solà (2001) and Yang *et al.* (1999). The ultimatum game has been studied extensively beginning with Güth *et al.* (1982); Roth (1995) reviews the literature.

dependence' over outcomes is captured by some of the social utility models based on the notion that fairness is driven by reciprocal-kindness (in Section 3 we will see that reciprocal-kindness is not critical to menu dependence).[11]

The third game in this study is the battle-of-the-sexes-with-fair-procedure (BOSFP, panel 3 of Figure 1). This game permits the BOS proposer the option of choosing an unbiased procedure – a lottery – to select the proposal. 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 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. (The hypothesis makes no immediate prediction for the comparison of UG and BOSFP with the basic BOS.) Note that, in contrast, the standard perfect equilibrium for all games in Figure 1 has the proposer offering the (1800, 200) split and the responder accepting.

Also note that menu dependence alone does not imply differing acceptability of outcomes across BOS and BOSFP since the menu of outcomes is fixed across games. On the other hand, menu dependence plus the first fair procedures hypothesis imply that rejection rates of the (1800, 200) proposal will be higher in BOSFP than in BOS, a point we will elaborate on in Section 3.

None of the social utility models speaks to the 'acceptability' of a procedure as such; random procedures enter decision making only through an evaluation of the involved risks. In particular, BOSFP and UG would be equivalent games for risk neutral and solely self-interested players. Risk-aversion reduces the attractiveness of the unbiased lottery compared to the equal split. Moreover, a risk-neutral but inequality-averse player would take into account that the fair procedure in BOSFP *always* yields unequal distributions. For example, distribution-based social utility models, based on the notion that fairness is based on inequality aversion, suggest that the value of the unbiased lottery in BOSFP is smaller than the equal split in UG, and that the lottery may even be rejected.[12] By these same models, the

[11] Falk and Fischbacher (1998), Cox and Friedman (2002), as well as Charness and Rabin (2002) suggest that (1800, 200) is more acceptable in BOS than in UG. Rabin (1993) and Dufwenberg and Kirchsteiger (forthcoming) also exhibit menu dependence but the kindness of the opponent's action in these models 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. Straightforward application 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. See Sen (1997) for a general discussion of 'menu dependence'.

[12] E.g., Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). Fehr and Schmidt's (1999, p. 822) utility function is linear and separable in absolute and relative payoffs, implying that players are risk neutral with respect to pecuniary payoffs. In their model the equal split in the UG has utility 1000, whereas the expected utility of the lottery in BOSFP is $1/2[200 - \alpha(1800 - 200)] + 1/2[1800 - \beta(1800 - 200)] < 1000$ for all $\alpha, \beta > 0$. While Bolton and Ockenfels' non-parametric model assumes neither linearity nor separability, they offer a linear, separable utility function (2000, p. 173) that assumes risk neutrality with respect to pecuniary payoffs (but non-linearity with respect to relative payoffs). By this, the utility of the equal split in the UG is $1000a$, whereas the expected utility of the lottery in BOSFP is $1/2[200a - (b/2)(1/2 - 200/2000)^2] + 1/2[1800a - (b/2)(1/2 - 1800/2000)^2] < 1000a$ for all $a, b > 0$. Thus, in both models, risk-neutral inequality-averse responders prefer the equal split over the unbiased lottery. Moreover, they would reject the lottery if the inequality-aversion is strong enough, that is if $\alpha + \beta$ or b/a, respectively, is sufficiently large.

acceptability of (1800, 200) is unaffected by the addition of the random procedure in BOSFP or the equal split in UG. This is because the relative payoff is a function of how the allocation compares to a reference point, typically a 50–50 allocation, and is independent of the 'outcome menu', a feature we modify when we discuss theory in Section 3.

1.1. *Laboratory Protocol*

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 to disclose the outcome of the first game until the second is complete. To simulate one-shot games more closely, two additional precautions are taken. After both games a coin flip picks just one game for actual payment. In this way subjects are effectively only matched with one partner in the experiment as a whole.[13]

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.[14,15] There is some empirical evidence that the form of play can make a difference (Schotter *et al.*, 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 behaviour, thereby dampening down treatment effects (at least so long as any effect of the strategy method is symmetric, but see Section 4).[16] The treatment effects behind the conclusions we draw, however, are all quite large (and in this sense the cool frame strengthens the results). When a conclusion rests on the absence of a treatment effect, the absence is quite unambiguous.

[13] A no-feedback design does not investigate the effect of experience. Previous studies of similar games to those here report no or small learning effects; e.g., Bolton and Zwick (1995), Kagel *et al.* (1996), Duffy and Feltovich (1999), Abbink *et al.* (2001), Bolton and Ockenfels (forthcoming *a*) and Cooper *et al.* (2003). 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, so a modest experience effect would not substantially alter our conclusions.

[14] This 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 3 proposals. See Section 2.

[15] One of the benchmark social utility models (Rabin, 1993) technically applies only to normal form games but see Dufwenberg and Kirchsteiger (forthcoming) for an extensive-form version.

[16] 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 *et al.* (2003) and Güth *et al.* (2001) find some support in the context of sequential two-player bargaining experiments.

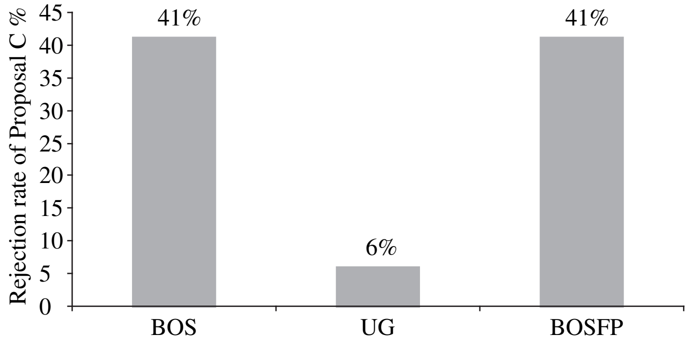


Fig. 2. *Rejection Rates of Proposal C in the Ultimatum and Battle-of-the-sexes Games*

A total of 204 people participated in the experiments reported in the article, with 102 participating in study 1. No subject participated in more than one session. The sample sizes are 34 for UG, 36 for BOS, and 32 for BOSFP, 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.

1.2. *Results from Study 1*

The complete data set is broken out in Appendix B both by player role and by individual subject. 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 difference in proportions test, $p < 0.001$ for each comparison separately).
- A fair procedure may substitute for a fair outcome: (*i*) the rejection rate of Proposal *C* is nearly identical in BOSFP (40.6%) and UG (41.2%); (*ii*) the unbiased proposal (Proposal *B*) and the unbiased procedure are both 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 favour 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 (though one might argue that it is harder to get confused over the unbiased proposal); in fact the latter is almost 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). At least 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 favourable situation (as it is also expressed, yet in a weaker form, in dictator games).[17]

Turning to proposer behaviour, we find that BOSFP proposers are more likely to choose Proposal *C* than *the other proposals* in comparison to UG proposers (56 vs. 29%; χ^2-test, p = 0.051). This cannot be justified by responder behaviour, because rejection rates are almost identical for each of the three proposals of the two games. Different explanations are conceivable. For one, both risk-averse and inequality-averse (as explained above) proposers prefer the fair allocation over the lottery between the biased allocations. Consequently, to the extent to which risk or inequality aversion matters, besides the procedural fairness aspect of the lottery, proposers might be more willing to risk rejection in the fair procedures game (later, we will speak in more detail to the heterogeneity of competing motives). It is also possible that BOSFP proposers (wrongly) anticipated lower rejection rates for Proposal *C* or higher rejection rates for Proposal *B* than UG proposers did. Our experiments in study 1 cannot disentangle these explanations for the change of proposer behaviour. However, any explanation based on risk attitudes, inequality aversion, or 'out-of-equilibrium' beliefs is not able to accommodate the data from our experiments in study 2, which are favourable to the relevance of procedural fairness.

Also, all proposers in the BOS treatment select Proposal *C*. In the other two games taken together, 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 *et al.*, 1998; Murnighan *et al.*, 2001): if a favoured moderate outcome is removed, people tend to move to an outcome that favours self relative to the moderate outcome. Before the experiments, we conjectured that individual behaviour as a proposer and as a responder would be correlated. Fehr and Schmidt's (1999) model implies that a proposer who chooses the fair offer is, as a responder, more likely to reject an unfair offer and *vice versa*. Although this is not the focus of our analysis, let us point out that some statistically significant patterns do relate proposer and responder behaviour (see the individual data in Appendix B). If one looks at the number of subjects that reject *C* in BOSFP, this occurred for 9 out of those 14 subjects that proposed *B*, but only for 4 out of those 18 that proposed *C*. This is significant with p = 0.016. In the UG, 12 out of 23 who chose *B* reject *C*, but only 2 out of 10 who chose *C*, p = 0.043 (one-sided).

[17] Several other experimental studies including Bornstein and Yaniv (1998), Güth *et al.* (2001), Hennig-Schmidt *et al.* (2001), Mitzkewitz and Nagel (1993) and Roth *et al.* (1991) also report rejections of offers higher than the equal split. Reviewing the literature and their own evidence, Hennig-Schmidt *et al.* (2001) conclude that 'Rejecting offers higher than the equal split appears to be a phenomenon occurring in different cultures and in student as well as in non-student populations.' From evaluating video experiments, they are able to attribute this finding to explicitly stated inequality aversion arguments; see also Güth *et al.* (2002) who come to a similar conclusion in a three-person bargaining context. As a referee pointed out, most social utility model reference points for fairness are inconsistent with rejecting offers higher than the equal split. However, Bolton and Ockenfels' (2000) model is consistent with this phenomenon as well as with related phenomena in centipede game behaviour (see p. 180).

To summarise the main results, we 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. However, even though responders treat allocation and procedural fairness identical, proposers do not, possibly reflecting concerns of risk or inequality aversion. Consistent with previous studies, we find that unfair outcomes are more acceptable if neither procedural nor outcome fairness is feasible.[18]

2. Study 2: Biased and Unbiased Randomisation in the Mini-ultimatum Game, *Ex post*

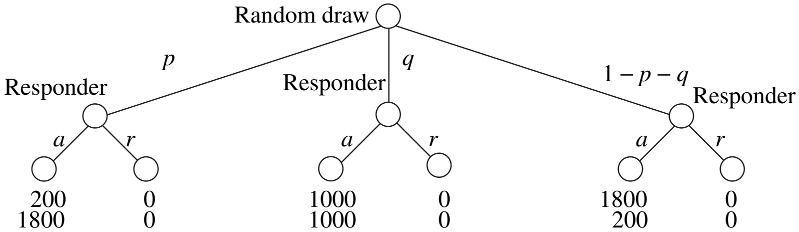
In study 2, we investigate whether the bias of the random procedure influences the *ex post* acceptability of the outcome of the procedure. The fair procedure hypothesis suggests that a biased allocation is more acceptable if chosen by an unbiased procedure than if chosen by a biased one. An alternative hypothesis, the attribution hypothesis, and a common interpretation of Blount's (1995) results, asserts that the important aspect of a random procedure is that a biased offer is no longer attributable to the proposer; so 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 two reasons. First, both the fair procedures and attribution hypothesis presuppose a substantial rate of rejection of biased offers in the non-randomisation 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 (5.6%), whereas the rejection rate for the ultimatum game is quite substantial (41%). Second is comparability; Blount's (1995) 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.[19] Minimally acceptable offers in the latter type of game were roughly half what they were in the standard game.

To the study 1 treatments 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 accepts or rejects the resulting draw. The payoff nodes are the same as for the UG. In the asymmetric treatment (ASYM) the probability distribution over proposals (A, B, C) is (p, q,

[18] A referee rightly points out that our role reversal design may somewhat lower the rejection rate of C in BOS compared to the UG because rejecting C either means rejecting the choice made as a proposer (if the responder chose C as a proposer), or it means rejecting the same allocation offered as a proposer (if the responder chose A as a proposer). In the UG, on the other hand, those who propose B do not face such a conflict. However, various earlier studies of UG and BOS found the same comparative effect (see footnote 10), and our arguments in this article do not depend on the size of this effect. We will come back to some design issues in our conclusions.

[19] 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%.



Treatments: (1) ASYM: $(p, q, 1 - p - q) = (0.01, 0.01, 0.98)$, (2) SYM 98: (0.01, 0.98, 0.01), and (3) SYM34: (0.33, 0.34, 0.33)

Fig. 3. *Ultimatum Games with Random Proposal*

$1 - p - q) = (0.01, 0.01, 0.98)$, in the first symmetric treatment (SYM98) the probability distribution is (0.01, 0.98, 0.01) and, in the second symmetric treatment (SYM34) we have (0.33, 0.34, 0.33). 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, indicating 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 the three new treatments.

As to the benchmark social utility models: Reciprocal-kindness models imply that the responder rejects an offer to punish the bad intentions of the proposer, consistent with the attribution hypothesis. Rejections of the unequal offer should then be similar across all three randomisation treatments but this common rate should be lower than that in the UG. In contrast, distribution-based models predict that rejection rates are determined by the division of the pie proposed, independent of how that division was arrived at. Rejection rates of Proposal *C* should then be similar across all randomisation treatments and the UG game. Note that risk preferences and 'out-of-equilibrium' beliefs cannot play a role in this game since the game involves a single choice made over certain outcomes.

2.1. *Laboratory Protocol*

With the exception of the method for selecting a proposal, all other procedures were the same as for study 1, including the use of the strategy method (Section 1.1). 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 experimenter 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.

2.2. *Results from Study 2*

The main treatment effect is displayed in Figure 4. (The complete data set is in Appendix B.) Since proposers make no selection in the new treatments, all observations concern responders:

- Proposals made by a fair random procedure lead to a different reaction than those made directly by a human player: Rejection rates for Proposal *C* in SYM34 and SYM98 are smaller than in the UG (one-sided chi-square independence or difference in proportions test, $p = 0.024$, for each comparison).
- The procedures that are unbiased have very similar effects: Rejection rates for Proposal *C* in SYM34 and SYM98 do not differ (the rejection rate for both is 18.8%).
- An unfair procedure is not a substitute for a fair outcome: There is no statistical difference in Proposal *C* rejection rates in UG and ASYM (two-sided $p = 0.542$).
- Rejection rates for Proposal *C* in ASYM are higher than in SYM34 or SYM98 ($p = 0.074$ for each comparison, one-sided; both hypotheses and data suggest it is permissible to pool the symmetric treatments, doing so yields one-sided $p = 0.040$).

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 (Section 1.2), Proposal *A* is rejected 10 times (about 10%) in total. There is no statistically significant difference, at any conventional level, between rejections rates across treatments for Proposal *B* respectively regardless of whether the data of the symmetric treatments are pooled or not. For Proposal *A*, there is also no significant difference, at any conventional level, when pooling the data. However, if one compares for this proposal the rejection rates in ASYM and in SYM34 one obtains $p = 0.039$ in a one-sided test.

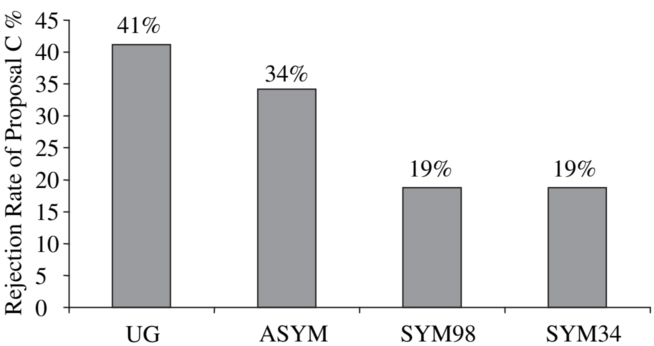


Fig. 4. *Rejection Rates of Proposal C in Ultimatum Games*

To summarise, the 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 randomisation is somewhat lower but 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.

3. Discussion. Towards an Explanation: the Relationship with Allocation Fairness

From study 2, we see that the rate of rejection of Proposal *C* varies with the selection procedure, holding potential allocations constant, controlling for risk and for the attribution hypothesis. This suggests that an explanation focused solely on allocation fairness, be it distributional or reciprocal-kindness forms of allocation fairness, will be unsatisfactory. We might then posit that the behaviour we observe, or at least that related to randomisation, requires a wholly new explanation separate from allocation fairness. But this 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 demonstrate that a combination of the social utility characterisation of allocation fairness with some simple precepts of procedural fairness is consistent with our data. We hasten to note that our intent is *not* any broader than this. Out intent is not, for instance, to introduce a new social utility model. The explanation we sketch here is *post hoc*, and leaves some important questions unanswered, questions that a full model of procedural fairness must address (we discuss some of these questions below).

3.1. *The Role of Reference Points in Models of Social Utility*

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. Likewise, adding random procedures may introduce fairness norms specific to procedural fairness, something suggested by the Rawls quotation at the beginning of the article. The reference points defined by distribution-based fairness models are generally invariant to the removal of the equal division offer, while some of the proposed measures of reciprocal-kindness predict an effect (Section 1). None of the models captures procedural fairness. While we think either sort of existing social utility model – distribution or reciprocal-kindness – can be extended to fit our data, we will focus here on extending a distribution-based model. One reason is that 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. 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 behaviour of the *experimenter*.[20] This may be 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 distribution or attribution based or both – will not capture this sort of procedure-sensitive behaviour so long as they only incorporate allocation-driven reference points.

3.2. *The Generalised Offer Game*

To put matters in a common framework, we first define a *generalised offer game*. There are two players: a proposer and a responder. The players jointly select a division of a pie normalised to size 1 from a finite and non-empty set, Ω, of feasible allocation plans. If $\omega \in \Omega$, then either ω is an efficient allocation (i.e., $\omega_1 + \omega_2 = 1$, subscripts denote which mover the share goes to), or ω 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 Λ; each $\lambda \in \Lambda$ specifies probability weights over the allocation plans in Ω. An offer of (1800, 200) is treated as a lottery giving weight 1 to (0.9, 0.1) and 0 to all other allocations (payoffs stated as shares of the pie). The choice of λ leads to the realisation of some ω; that is, uncertainty over Ω specified by λ is resolved. In stage 2, the responder either accepts or rejects the resulting ω. If accepted, ω is realised 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 (Λ) and the set of allocation plans (Ω) 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 (0.9, 0.1), one assigning weight 1 to (0.1, 0.9) and one assigning weight 1/2 to each possible allocation. The responder directly accepts or rejects the proposer's choice. So for this game, $\Omega = \Lambda$. For SYM34, the proposer chooses from a singleton set (the fair lottery). After the procedure is realised, the responder accepts or rejects the selected allocation, meaning $\Omega =$ (0.9, 0.1), (0.5, 0.5), (0.1, 0.9). The framework also accommodates the ultimatum game: $\Lambda = \Omega =$ (0.9, 0.1), (0.5, 0.5), (0.1, 0.9). For BOS, $\Lambda = \Omega =$ (0.9, 0.1), (0.1, 0.9).[21]

[20] 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 which finds that responders are rather insensitive to who makes the offer (also see Section 4 of the present article). This explanation contradicts the definition of the reference *group* given in the economic theories of reciprocal-kindness (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 summarised 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.

[21] Strictly speaking, in UG, BOS and BOSFP it is not precisely true that $\Lambda = \Omega$. Rather, the former is the set of lotteries over the elements of Ω that assign probability 1 to one of these elements. In relation to our study 1 we have argued that the availability of a fair *procedure* can replace the availability of a fair *allocation*. A referee has pointed out that it would be more precise to say that the availability of a procedure which with certainty selects a fair allocation plan can replace a fair allocation. Although equating the set of procedures and the set of allocation plans does somewhat blur the sequence of moves, we feel that the use of the term *procedure* in the early parts of the article better conveys the basic motivation behind what we do.

3.3. *A Model*

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

$$v_i = v_i(\omega_i, \sigma_i).$$

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

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

If σ_i is measured as

$$\sigma_i = \sigma_i(\omega_i^e) = \begin{cases} \omega_i^e, & \text{if } \omega \in \Omega \\ 1/2, & \text{if } \omega = (0,0), \end{cases} \tag{1}$$

we have the ERC model proposed by Bolton and Ockenfels (2000). The only difference is that (1) accounts for the fact that an outcome may be a lottery by evaluating the relative payoff at its expected value ω_i^e. 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 maximises everyone's relative payoff. This division can be thought of as a socially recognised fairness norm.[22]

Formulation (1) of the relative payoff is consistent with regularities observed over a variety of games (Bolton and Ockenfels, 2000). 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; that is the reference point of fairness is menu-dependent (see Section 1), as it is in most reciprocal-kindness models. An obvious choice would be the feasible outcome closest to equal division. Let $m_d = \min_\Omega |\omega_i^e - 1/2|$. For $m_d < 1/2$, define

$$\sigma_i' = \sigma_i'(\omega_i^e, m_d) = \begin{cases} \omega_i^e/(1-2m_d), & \text{if } \omega_i^e \leq 1/2, \omega \neq (0,0) \\ (\omega_i^e - 2m_d)/(1-2m_d), & \text{if } \omega_i^e > 1/2, \omega \neq (0,0) \\ 1/2, & \text{if } \omega = (0,0). \end{cases} \tag{2}$$

Since $m_d = 1/2$ implies $\Omega \subseteq \{(1,0),\ (0,1)\}$, we define: $\sigma_i'(\omega^e, 1/2) \equiv 1/2$. Equation (2) is a piece-wise linear rescaling of (1), so that σ_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).[23] By formulation (2), in BOS both (1800, 200) and (200, 1800) are

[22] 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.

[23] In particular, this implies that for all $\omega \in \Omega$ and all $m_d < 1/2$: $\sigma'_i(1, m_d) = 1$, $\sigma_i'(0, m_d) = 0$, $\sigma_i'(1/2 - m_d, m_d) = 1/2$, $\sigma_i'(1/2 + m_d, m_d) = 1/2$, and $\sigma_i'(\omega^e, 0) \equiv \omega_i^e$ for $\omega^e \neq (0,0)$.

recognised as fair outcomes (both are the same minimum distance from equal division), whereas by formulation (1) neither is recognised as such.[24]

To add consideration of procedural fairness to (2), let λ_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 λ_i^e or σ_i' depending on whether the outcome or the procedure is less biased, i.e., which is closer to 1/2:

$$\sigma_i'' = \sigma_i''(\omega_i^e, m_d, \lambda_i^e) = \arg\min_{\sigma_i'(\omega_i^e, m_d), \lambda_i^e}[|x - 1/2|]. \tag{3}$$

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, λ_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.

3.4. *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 $\omega = (0.9, 0.1)$. Then $\sigma_i(0.1) = \sigma_i'(0.1, 0) = 0.1$, but since $\lambda_i^e = 1/2$ we have $\sigma_i''(0.1, 0, 1/2) = 1/2$, and so $v_i = v_i(0.1,1/2)$. The value of rejecting is always $v_i = v_i(0,1/2) < v_i = v_i(0.1,1/2)$ implying the offer will be accepted independent of i. Now, suppose that the offer of (0.9, 0.1) is generated by the ASYM random procedure. Again $\sigma_i(0.1) = \sigma_i'(0.1, 0) = 0.1$, but now $\lambda_i^e = 0.11$ (computed using the ASYM probability weights) implying the offer may be rejected. On the other hand, by (3) a fair outcome (0.5, 0.5) is never rejected, regardless of the procedure.[25]

Also, for BOSFP, offering (0.9, 0.1) (with probability 1) implies $\sigma_i(0.1) = \sigma_i'(0.1, 0) = 0.1$, $\lambda_i^e = 0.1$ and so $v_i = v_i(0.1,.1)$, the same value as for the standard ultimatum game, and, as in the latter case, the proposal may be rejected. In contrast, accepting the proposal to randomise in BOSFP has a value of $v_i = v_i(\rho_i,1/2)$ where ρ_i denotes the absolute payoff certainty equivalent of the randomisation to i. Given that the worst absolute payoff that might result from the lottery is 0.1, it is safe to assume that $\rho_i > 0$, making $v_i(\rho_i,1/2) > v_i(0,1/2)$; the proposal should be accepted (note, however, that if the certainty equivalent of the fair lottery is not 1/2, then $m_d > 0$).

[24] 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-centred perception of fairness; e.g., a player may only care how fair of a share is feasible for himself 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

[25] In line with this prediction, the fair outcome is rejected in less than 3% of all cases. Three out of the 4 rejections, however, occur in ASYM and are combined with rejections of A but not of C. Thus, this might be attributable to the low probability. In this sense, one may see this feature of the data as an unintended but unavoidable consequence of the use the strategy method, which, as pointed out before, has a number of advantages that speak in favour of its use.

If we suppose that all players gauge the fairness of an outcome by (3), we can explain the *comparative statics* 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).

To get a further feel for the implications of the model, consider a situation not covered by our experiments. What would happen if in ASYM the probability of Proposal *C* were 100%? The model tells us to expect a similar rejection reaction as in the ASYM treatment with 98% that we report on in the article. The prescription is hence the same as for the case where a human proposer effectively chooses Proposal *C* (the UG treatment) and different from the BOS and a situation where only Proposal *C* existed. A natural interpretation of this involves issues of framing; we discuss this further in our concluding Section.

Some rejections in BOS and in the unbiased random offer games are not captured by reference point (3). A full 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).[26]

The main insight here is that competing fairness norms, appropriately embedded in reference points, can explain much of the seemingly anomalous behaviour we observe in our experiments on fair procedures. Specifically, the model captures how a fair procedure can make biased outcomes less or more acceptable, depending on the circumstance. A fully developed model requires a still more refined statement of how people choose between competing fairness norms. To give an example, consider a game in which Ω is as in BOS. The model presently implies that the responder will accept (1800, 200) independent of whether the proposer chooses it by a coin flip or a lottery that puts weight 0.9 on (1800, 200) and (200, 1800) otherwise. That seems implausible. Another somewhat questionable prediction arises for the case in which in the BOSFP game the responder made his decision after playing out the lottery. The model predicts

[26] Other experimental evidence suggests that people differ in what they perceive as fair when there are multiple fairness norms to choose from; e.g., Roth and Malouf's (1979) early studies of Nash-bargaining, Babcock *et al.*'s (1995) and Hennig-Schmidt's (2000) study of the self-serving bias in fairness judgements, and Ockenfels and Weimann's (1999) cross-culture study.

the same rejection rates for the unfair allocation chosen with certainty in this game and in the BOS. However, given the option of having a lottery determine who obtains the lion's share one would, intuitively, expect more rejections in this game than in the BOS. The problem is that the model presently offers no credible decision rule of what is fair when there is more than one lottery to choose from.

Another detail of our data that can not be fully accommodated by our model is the difference in the rejection rates of Proposal *A* between treatments ASYM and SYM34. Here the model should lead to the expectation of fewer rejections under ASYM, since a favourable outcome resulting from an unfavourable lottery should approximately cancel each other out. But perhaps we should revert here to the spirit of our explanation and not apply the model mechanically. We find that a biased allocation is more acceptable when it results from a fair procedure than from an unfair one. Surprisingly, this notion appears to have some force even when the allocation is biased in one's favour. One may, however, judge this to be of somewhat lesser practical importance, since with ASYM the proportion of realisations that lead to *A* is small.

4. 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, centred as they are on unbiasedness, speak to a general phenomenon.[27] Our data indicate that choice behaviour is sensitive to procedural fairness: 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, though *random* fair procedures also add risk components to the situation that may also affect behaviour. Results produced by an unbiased procedure tend to be more acceptable than those produced by a biased procedure. These findings begin to explain the social pressure to implement procedures deemed fair. Simply put, some people are willing to impose a cost – both on self and others – to resist procedures that they deem to be biased against them.[28]

[27] While the data from this experiment are consistent with the notion that people have a preference for egalitarianism (a strong version of relative payoff preferences), they are also consistent with a more self-centred 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-centred concern for equity. While Fehr and Schmidt (1999) assume that players have egalitarian preferences, Bolton and Ockenfels (2000) use a 'self-centred fairness' formulation that presumes that players only care about their own relative status. The evidence in games with more than two players – where the difference has bite – is mixed; see the discussion and references in Bolton and Ockenfels (forthcoming *b*).

[28] Greenberg (1990) reports a natural field 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.

Our experiment abstracts away from other aspects of the field procedures – for example, physical versus (experimenter) mandated indivisibility – which one might conjecture are critical to people's judgment that randomisation is fair. But in fact, we observe strong reactions to the randomisation in the same direction anecdotal evidence suggests, implying that these other factors are not decisive.

We argued in the last Section that the results of our experiment require an explanation that meshes procedural fairness with allocation fairness, and demonstrated how such an approach might work by sketching an extension of a social utility model that includes competing fairness norms. Binmore *et al.* (2002, p. 85) 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. Our model sketch suggests that the search for a systematic manipulation of preferences might fruitfully be narrowed to a search for a systematic manipulation of the reference point. Different classes of games might systematically evoke different reference points (fairness norms), giving rise to a practical taxonomy.

The model we sketch is incomplete in several ways. For one, it 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.[29] While other modelling approaches are certainly possible, 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 behaviour independent of the game, any model with a simple statement of those norms is bound to be incomplete. Such models can nevertheless be useful: work to date has demonstrated that we can go a surprisingly long way, in some important classes of games, with very simple models of fairness. For more challenging games, as shown in Section 3, systematic variations of fairness norms can be incorporated into these models in a way that allows subsequent empirical tests and theoretical refinements.

Our model implies that, in judging fairness, people focus more on some aspects of the situation they find themselves in than on others. If they had focused more on the fact that roles were assigned randomly (see Appendix A), they might have concluded that a biased allocation from an ASYM draw is as fair as one from a SYM

[29] Frey (1999) exhibits a more complex scenario that includes a random procedure different from the kind studied here. He reports that respondents to a written survey, studying the fairness of alternative allocation procedures, largely refuted a 'random procedure (e.g., to give all persons whose surname starts with A to P)' as a means of distributing excess-demanded water among hikers at a point reachable only on foot. Most people preferred the 'first come, first served' procedure, which Frey terms the 'traditional' procedure. As noted earlier, a key feature of a fair procedure is that it levels the playing field, and 'first come, first serve' can be interpreted in this way, and in fact is arguably fairer in this sense than a rule that gives priority according to surnames. Some models in industrial organisation equate first come, first serve with the kind of random selection we study here. One might also favour first come, first serve for incentive reasons: Respondents might have reasoned that those hikers who need the water most are more likely to make the effort necessary to arrive first (efficient rationing). Efficiency concerns are excluded from our study, since all proposed allocations are efficient.

draw (most outcomes in our experiment could be rationalised as 'fair' in this way). Yet the data firmly reject this. Most subjects apparently take into account *narrow* procedural aspects of the games (i.e., excluding role assignment rules). This may simply be an inadvertent consequence of how the experimental situation was framed to subjects. On the other hand, it may be indicative of a natural propensity, when considering what is fair, to look back a finite number of steps; it seems unlikely to us that people look all the way back to their birth when evaluating the fairness of most situations.[30]

Our study employed the strategy method for all treatments. This gave us comparability with Blount's study and is also perhaps the only practical way to get statistically meaningful data on small probability events (e.g., the rejection of offers made but occasionally). So long as the effects of the strategy method are symmetric on all treatments, we would expect our main conclusions to hold for the non-strategy method case: seeing the first mover's actual offer before deciding, might lead to a higher probability of rejection (perhaps due to greater emotion) but this would be true for all treatments; hence the base level of rejection would change but not the comparative statics. Alternatively, one might conjecture that the strategy method effects are asymmetric. When the strategy method is employed the cognitive focus may be on the procedure, so that subjects are willing to commit to unfair outcomes if they result from a fair procedure, while if they played without the strategy method, then the cognitive focus would be on the outcome and hence they would be more likely to reject an unfair outcome even if it results from a fair procedure. The important implication would be that the difference between ASYM and SYM rejection rates might disappear in a non-strategy method implementation. A referee pointed this possibility out to us but also noted (and we concur) that it seems unlikely, given that other kinds of menu dependence effects have been demonstrated in non-strategy method designs. Existing evidence on differences between the strategy method and the direct response mode is mixed. At the same time, there is evidence that behaviour may be menu-dependent, even if the strategy method is not used (Brandts and Solà, 2001; Bolton and Ockenfels, forthcoming *a*; Falk *et al.*, 2003). So while it is possible that a non-strategy method implementation of our games would change the results, the present evidence does not directly imply it. At any rate, a specific investigation of this issue is beyond the scope of this article.

Finally, our experiment raises an interesting 'credibility' issue. Think of a seaman claiming to have randomly selected the passenger to be thrown overboard, 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 *observable* proof that these decisions were arrived at by a fair procedure, it is hard to imagine the disaffected – doomed passengers, alienated citizens, or losing responders – accepting it. Only observable and verifiable

[30] The question is worthy of further investigation. One working hypothesis is that people reason back myopically, just as they do forward; see, e.g., Selten and Stöcker (1986) and Johnson *et al.* (2001). More generally, issues of perception may be quite central. If we changed the game to bring all procedures to subjects' attention, then they might think the ASYM procedure more fair; Blount (1995) has a related framing effect.

unbiased procedures would seem *credibly* fair.[31] A firm conclusion requires further investigation.

Penn State University
Institut d'Anàlisi Econòmica (CSIC), Barclona
University of Cologne

Submitted: 8 January 2002
Accepted: 25 August 2004

A Technical Appendix is available for this paper at http://www.res.org.uk/economic/ta/tahome.asp

References

Abbink, K., Bolton, G.E, Sadrieh, K. and Tang, F. (2001). 'Adaptive learning versus punishment in ultimatum bargaining', *Games and Economic Behavior*, vol. 37(1), pp. 1–25.

Andreoni, J., Brown, P.M., Vesterlund, L. (2002). 'What makes an allocation fair? Some experimental evidence', *Games and Economic Behavior*, vol. 40, pp. 1–24.

Anand, P. (2001). 'Procedural fairness in economic and social choice: evidence from a survey of voters', *Journal of Economic Psychology*, vol. 22, pp. 247–70.

Babcock, L., Loewenstein, G., Issacharoff, S., and Camerer, C. (1995). 'Biased judgements of fairness in bargaining', *American Economic Review*, vol. 85(5), pp. 1337–43.

Binmore, K., McCarthy, J., Ponti, G., Samuelson, L. and Shaked, A. (2002). 'A backward induction experiment', *Journal of Economic Theory*, vol. 104(1), pp. 48–88.

Blount, S. (1995). 'When social outcomes aren't fair: the of causal attributions on preferences', *Organizational Behavior and Human Decision Processes*, vol. 63, pp. 131–44.

Bolton, G.E (1991). 'A comparative model of bargaining: effect theory and evidence', *American Economic Review*, vol. 81, pp. 1096–136.

Bolton, G.E, Katok, E. and Zwick, R. (1998). 'Dictator game giving: rules of fairness versus acts of kindness', *International Journal of Game Theory*, vol. 27, pp. 269–99.

Bolton, G.E, and Ockenfels, A. (2000). 'ERC: a theory of equity, reciprocity and competition', *American Economic Review*, vol. 90(1), pp. 166–93.

Bolton, G.E, and Ockenfels, A. (forthcoming *a*). 'A stress test of fairness measures in models of social utility', *Economic Theory*.

Bolton, G.E, and Ockenfels, A. (forthcoming *b*). 'Self-centered fairness in games with more than two players', in (C. Plott and V. 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*, vol. 10, pp. 95–121.

Bornstein, G., and Yaniv, I. (1998). 'Individual and group behavior in the ultimatum game: are groups more rational players?', *Experimental Economics*, vol. 1, pp. 101–8.

Brandts, J., and Charness, G. (2000). 'Hot vs. cold: sequential responses and preference stability in experimental games', *Experimental Economics*, vol. 2, pp. 227–38.

Brandts, J., and Solà, C. (2001). 'Reference points and negative reciprocity in simple sequential games', *Games and Economic Behavior*, vol. 36(2), pp. 138–57.

Brosig, J., Weimann, J., and Yang, C. (2003). 'The hot versus cold effect in a simple bargaining experiment', *Experimental Economics*, vol. 6, pp. 75–90.

[31] 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 found little substantial difference in rejection rates between the standard ultimatum and third party treatments: '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 randomisation.

Charness, G. and Rabin, M. (2002). 'Understanding social preferences with simple tests', *Quarterly Journal of Economics*, vol. 117, pp. 817–69.

Cooper, D.J., Feltovich, N., Roth, A.E. and Zwick, R. (2003). 'Relative versus absolute speed of adjustment in strategic environments: responder behavior in ultimatum games', *Experimental Economics*, vol. 6, pp. 181–207.

Cox, J.C., and Friedman, D. (2002). 'A tractable model of reciprocity and fairness'. Working Paper, University of Arizona.

Duffy, J. and Feltovich, N. (1999). 'Does observation of others affect learning in strategic environments? An experimental study', *International Journal of Game Theory*, vol. 28(1), pp. 131–52.

Dufwenberg, M., and Kirchsteiger, G. (forthcoming). 'A theory of sequential reciprocity', *Games and Economic Behavior*.

Elster, J. (1989). *Solomonic Judgements*, Cambridge: Cambridge University Press.

Falk, A., Fehr, E., and Fischbacher, U. (2001). 'Informal sanctions'. Working Paper, University of Zürich.

Falk, A., Fehr, E. and Fischbacher, U. (2003). 'On the nature of fair behavior', *Economic Inquiry*, vol. 41, pp. 20–6.

Falk, A., and Fischbacher, U. (1998). 'A theory of reciprocity'. Working Paper, University of Zurich.

Fehr, E., and Schmidt, K. (1999). 'A theory of fairness, competition, and cooperation', *Quarterly Journal of Economics*, vol. 114, pp. 817–68.

Frey, B.S. (1999). *Economics as a Science of Human Behaviour: Towards a New Social Science Paradigm*, Boston: Kluwer Academic Publishers.

Frey, B.S., and Stutzer, A. (2001). 'Beyond Bentham: measuring procedural utility', University of Zurich working paper.

Frey, B.S., Benz, M. and Stutzer, A. (2002). 'Introducing procedural utility: not only what, but also how matters', Working Paper No. 129, Institute for Empirical Research in Economics, University of Zürich.

Forsythe, R., Horowitz, J.L., Savin, N.E., and Sefton, M. (1994). 'Fairness in simple bargaining experiments', *Games and Economic Behavior*, vol. 6, pp. 347–69.

Greenberg, J. (1990). 'Employee theft as a reaction to underpayment inequity: the hidden cost of pay cuts', *Journal of Applied Psychology*, vol. 75, pp. 561–8.

Güth W., Huck, S., and Müller, W. (2001). 'The relevance of equal splits in ultimatum games', *Games and Economic Behavior*, vol. 37, pp. 161–9.

Güth, W., Schmidt, C., and Sutter, M. (2002). 'Bargaining outside the lab - a newspaper experiment of a three-person ultimatum game', working paper, Max Planck Institute, Jena.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). 'An experimental analysis of ultimatum bargaining', *Journal of Economic Behavior and Organization*, vol. 3, pp. 367–88.

Hennig-Schmidt, H. (2000). 'Perceptions of fairness - an analysis of different video experiments', in (E. Hölzl, ed.), Conference Proceedings of IAREP/SABE 2000, Vienna: WUV-Universitätsverlag, pp. 194–7.

Hennig-Schmidt, H., Li, Z.-Y., Yang, C. (2001). 'Non-monotone strategies in ultimatum bargaining: first results from a video experiment in the people's Republic of China', *Proceedings of the International Congress of Mathematicians Game Theory and Applications Satellite Conference*, Qingdao Publishing House, pp. 225–31.

Johnson, E.J., Camerer, C., Sen, S., and Rymon, T. (2001). 'Detecting failures of backward induction: monitoring information search in sequential bargaining', *Journal of Economic Theory*, vol. 104, pp. 16–47.

Kagel, J., Kim, C., and Moser, D. (1996). 'Fairness in ultimatum games with asymmetric information and asymmetric payoffs', *Games and Economic Behavior*, vol. 13, pp. 100–10.

Kagel, J. and Roth, A.E. eds. (1995). *Handbook of Experimental Economics*, Princeton: University of Princeton Press.

Kunreuther, H., and Portney, P. (1991). 'Wheel of fortune: a lottery/auction mechanism for siting of noxious facilities', *Journal of Energy Engineering*, vol. 117, pp. 125–32.

Lind, E.A., and Tyler, T.R. (1988). *The Social Psychology of Procedural Justice*, New York: Plenum Press.

Mitzkewitz, M., and Nagel, R. (1993). 'Envy, greed and anticipation in ultimatum games with incomplete information', *International Journal of Game Theory*, vol. 22, pp. 171–98.

Murnighan, K., Oesch, J.M., and Pillutla, M. (2001). 'Player types and self impression management in dictator games: two experiments', *Games and Economic Behavior*, vol. 37(2), pp. 388–414.

Oberholzer-Gee, F., Bohnet, I., and Frey, B.S. (1997). 'Fairness and competence in democratic decisions', *Public Choice*, vol. 91, pp. 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*, vol. 71(2), pp. 275–87.

Rabin, M. (1993). 'Incorporating fairness into game theory and economics', *American Economic Review*, vol. 83, pp. 1281–302.

Rawls, J. (1971). *A Theory of Justice*, Cambridge, MA: Belknap Harvard Press, (quote at the head of this article is 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*, pp. 253–348. Princeton: Princeton University Press.

Roth, A.E., and Malouf, W.K. (1979). 'Game-theoretic models and the role of information in bargaining', *Psychological Review*, vol. 86, pp. 574–94.

Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M. and Zamir, S. (1991). 'Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study', *American Economic Review*, vol. 81, pp. 1068–95.

Schotter, A., Weigelt, K., and Wilson, C. (1994). 'A laboratory investigation of multiperson rationality and presentation effects', *Games and Economic Behavior*, vol. 6, pp. 445–68.

Selten, R., and Stöcker, R. (1986). 'End behavior in sequences of finite prisoner's dilemma supergames', *Journal of Economic Behavior and Organization*, vol. 7, pp. 47–70.

Sen, A. (1997). 'Maximization and the act of choice', *Econometrica*, vol. 65, pp. 745–79.

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*, vol. 57, pp. 830–8.

Tyler, T.R., and Lind, E.A. (2000). 'Procedural justice', in (J. Sanders and V.L. Hamilton, eds.), *Handbook of Justice Research in Law*, pp. 63–91. New York: Kluwer.

Yang, C., Mitropoulos, A., and Weimann, J. (1999). 'An experiment on bargaining power in simple sequential games'. Working Paper, University of Magdeburg.