Centre for Behavioural and Experimental Social Science

University of East Anglia

2017-08R

DOES PROCEDURAL FAIRNESS MAKE INEQUALITY MORE ACCEPTABLE?

Robert Sugden

Mengjie Wang

www.uea.ac.uk/cbess

Does procedural fairness make inequality more acceptable?

Robert Sugden and Mengjie Wang

School of Economics and Centre for Behavioural and Experimental Social Science

University of East Anglia

Norwich NR4 7TJ

United Kingdom

email: r.sugden@uea.ac.uk and m.wang3@uea.ac.uk

Abstract: In many real world situations, unfairness of outcomes is not directly related to fairness-relevant properties of individual decisions; it is an unintended consequence of procedures in which individuals interact. Attitudes to such unfairness may be revealed in emotions of anger and resentment rather than in 'social preferences' over alternative decision outcomes. We conjecture that inequality is viewed with relatively little disfavour when it results from procedures that allow individuals equal opportunities to compete. We define a concept of procedural fairness which formalises intuitions about equality of opportunity. We report a 'vendetta game' experiment in which emotions triggered by procedural unfairness can induce costly and counter-productive 'taking' of co-players' assets. A given degree of material inequality induces more taking if the procedure that has generated it is unfair rather than fair. Surprisingly, there is excess taking by players whom procedural unfairness has benefited as well as by those it has harmed.

Keywords: procedural fairness; inequality; vendetta game

JEL classifications: C92, D63, D91

Acknowledgements: Financial support for the experiment was provided by the Centre for Behavioural and Experimental Social Science (CBESS) and the University of East Anglia. Sugden's work on this project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 670103. We thank Bertil Tungodden and Daniel Zizzo for advice and comments.

Does procedural fairness make inequality more acceptable?

1 Introduction

Contributors to the social preference literature sometimes point to an apparent contrast between behaviour in two classes of experiment. Experiments using dictator, ultimatum, trust and public good games have found evidence of many types of other-regarding motivations, such as preferences for increasing social welfare, reducing inequality, rewarding kind intentions, punishing unkind intentions, rewarding effort, and respecting individuals' prior entitlements. In contrast, behaviour in market experiments is often well explained by the hypothesis of self-interest. When commenting on this difference, several writers have cited an experiment by Roth et al. (1991) that studied repeated induced-value markets in which nine buyers competed to buy one unit of a good from a single seller. Buyers' offers converged to the equilibrium price, at which the whole of the gains from trade were appropriated by the seller. In separate discussions of this result, Levine (1998: 605–606), Fehr and Schmidt (1999: 830) and Falk and Fischbacher (2006: 307) all offer the same explanation of buyers' willingness to accept this unequal outcome. In Roth et al.'s experiment, it is said, an individual buyer who proposes an equal division of the surplus will fail to trade, but the seller will trade with someone else at the high price. The buyer who insists on an equal share of the surplus loses out, but the outcome is still unequal, and the seller is not punished for demanding a disproportionate share. This example illustrates a common feature of market experiments, namely that participants are not able to bring about fair outcomes through individual action. To put this another way, the rules of the market are such that the fairness or unfairness of outcomes does not correspond in any direct way with the actions of individual participants.

This property is not peculiar to market experiments, or to markets. It occurs because the market is a procedure in which interactions between individuals can produce consequences that were not deliberately chosen by any individual participant. The concepts of fairness that feature in the social-preference literature are necessarily silent about the fairness or unfairness of non-chosen consequences. A social preference, as this term is normally understood, describes a person's attitude to choice problems which, at least in principle, she could face *as an individual*. But if an unequal outcome is the consequence of a complex interaction between independently-motivated individuals, it may have no direct relationship to any individual's decision problem.

This line of thought prompts the following question: What attitudes do people take towards inequality when no one has chosen it, but it is the outcome of a rule-governed interaction in which they have participated together with others? Do they see it as an undesired unfairness that unfortunately they are unable to do anything about? (That seems to be the hypothesis favoured by Levine, Fehr and Schmidt and by Falk and Fischbacher in the case of market experiments.) Or do they think that inequality is unfair only if it has been deliberately chosen? Our paper explores these questions. Specifically, we investigate a conjecture made by Isoni et al. (2014) when discussing evidence from experimental bargaining games – the conjecture that inequality is viewed with relatively little disfavour when it is the result of self-interested behaviour in an interaction in which individuals had equal opportunities to compete.

The issues we are addressing are recurring themes in the literature of economics. Philosophical economists who defend the market sometimes urge their readers to evaluate it only in terms of its general rules, while acknowledging that markets can produce outcomes which, had they been consciously chosen by any identifiable person, would be judged to be unfair. For example, Hayek (1976) poses the rhetorical question:

> Are we not all constantly disquieted by watching how unjustly life treats different people and by seeing the deserving suffer and the undeserving prosper? And do we not all have a sense of fitness, and watch it with satisfaction, when we recognize a reward to be appropriate to effort or sacrifice?' (p. 68)

Hayek accepts that the market can treat people unjustly in the same sense that life can: '[I]n the cosmos of the market we all constantly receive benefits which we have not deserved in any moral sense' (p. 94). But such perceptions of deservingness and undeservingness, although psychologically natural, are not relevant for an evaluation of the market:

> In a spontaneous order the position of each individual is the resultant of the actions of many other individuals, and nobody has the responsibility or the power to assure that these separate actions of many will produce a particular result for a certain person... There can, in a spontaneous order, be no rules which will determine what anyone's position ought to be. (p. 33)

Buchanan (1964, p. 219) voices a similar thought:

> The 'market' or market organization is not a *means* toward the accomplishment of anything. It is, instead, the institutional embodiment of the voluntary exchange processes that are entered into by individuals in their several capacities. That is all there is to it.

Nevertheless, these arguments are clearly intended as responses to a common belief that markets often *are* unfair. For Hayek and Buchanan, people's willingness to follow the rules of the market is a valuable form of social capital which that belief might erode. Rawls (1971, pp. 16, 177, 453–462) makes this point more generally when he argues that a social institution is more likely to sustain itself over time if it is *psychologically stable* – that is, if the operation of its rules tends to reproduce a general belief that those rules are fair.

If economics is to be able to say anything about psychological stability, it needs an analysis of *procedural* fairness, parallel with but distinct from the analysis of fairness in individual choice. Such an analysis needs an inventory of characteristics of procedural fairness, parallel with the fairness characteristics that feature in the literature of social preferences, and it needs to be able to elicit individuals' attitudes to different mixes of these characteristics. However, because of the disconnect between individual actions and outcomes, the elicitation of attitudes to procedural fairness is not straightforward.

One possible experimental approach would be to present an individual with a choice between alternative procedures to govern some interaction within some group of people. Such a decision task would be the analogue in the domain of procedures of a Dictator Game in the domain of outcomes: in effect, the respondent would be taking on the role of a social planner, choosing between institutional designs. But if one is concerned about the stability of institutions, one might be more interested in finding an analogue of the role of the second mover in an Ultimatum Game – that is, in finding situations in which a person can express her personal resistance to a procedure that she perceives as unfair. An Ultimatum Game second mover who rejects a disadvantageous offer can do so with the intention of punishing the first mover for taking an unfair action, or of preventing the first mover from gaining relative to her. But someone who is disadvantaged by the rules of a more complex procedure may be unable to identify any particular individual who is responsible for that disadvantage, and may have no way of reducing inequality between herself and anyone whose payoff is greater than hers. Perceptions of procedural unfairness might then reveal themselves in emotions of inchoate dissatisfaction, resentment or anger rather than in purposeful decision-making.

In this paper, we take a first step towards characterising procedural fairness, by defining a concept of procedural fairness which formalises the intuitive idea of equality of opportunity in a game-theoretic framework. In a first step of a different kind, we develop an experimental design in which negative emotions triggered by perceptions of procedural unfairness can be revealed in decisions that have material consequences. Using this design, we investigate whether a given degree of inequality between individuals' material outcomes is more acceptable if the procedure that has generated it is procedurally fair than if it is not.

The paper proceeds as follows. In Section 2, we briefly review the existing theoretical and experimental literature on fairness, contrasting procedural and non-procedural principles. In Section 3, we present our formal concepts of procedural fairness and bias. We then describe our experimental design (in Section 4) and the hypotheses it is intended to test (in Section 5). We report our results in Section 6 and discuss their implications in the final Section 7.

2 Procedural and non-procedural principles of fairness

In the theoretical and experimental literature about social preferences, principles of fairness are most commonly understood as properties of an individual's preferences over alternative allocations of material payoff between two or more individuals (possibly but not necessarily including the individual who holds the preferences). Many of the principles considered in this literature make no reference to the process by which the available surplus came into existence. Such principles include inequality aversion (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), the maximisation of social welfare or efficiency (e.g. Andreoni and Miller, 2002; Charness and Rabin, 2002), and kindness in the sense of being willing to forgo one's own payoff to achieve an 'equitable' allocation (e.g. Rabin, 1993). Other principles that have been studied require that individuals should be rewarded according to effort exerted, skill shown, contribution made or luck experienced in the creation of the surplus that is to be distributed (e.g. Ruffle, 1998; Konow, 2000; Cappelen et al., 2007; Capellen et al., 2013, Mollerstrom et al., 2015). Significantly, however, most of these studies (including all those just cited) use designs in which the surplus-creating process does not involve interaction between individuals: the surplus is simply the sum of the results of individuals' separate activities. An experimental set-up of this kind elicits the preferences of participants who have been assigned the role of a dictator or social planner who chooses how to reward individuals for actions they have already taken, or for characteristics they have already exhibited. The social preference literature also studies second-order principles of reciprocity. Such principles are understood as preferences, held by individuals, for rewarding other people for acting on principles of fairness and punishing other people for acting contrary to such principles (e.g. Rabin, 1993; Charness and Rabin, 2002; McCabe et al., 2003).

Notice that all these first- and second-order principles of fairness are structured to apply to cases in which a single decision-maker chooses between alternative payoff allocations. In this sense, they are non-procedural. They presuppose what Hayek argues cannot be true of a spontaneous order – that there can be rules that determine what each individual's outcome ought to be, independent of the workings of the order itself.

The social-preference literature we have considered so far is complemented by a much smaller literature about procedural fairness. However, the latter is mainly concerned with a very special form of such fairness, namely randomness. For example, Blount (1995) studies a variant of the Ultimatum Game in which the proposed payoff allocation is generated by an unbiased random device. She finds that second movers are more willing to accept disadvantageous inequality if the proposal is generated in this way than if it is deliberately chosen by a co-player. Bolton et al. (2005) investigate another variant of the Ultimatum Game in which the first mover chooses between three options – a proposal that favours him, a proposal that favours his co-player, and using an unbiased random device to select one of those two proposals. Second movers were more willing to accept disadvantageous proposals that had come about through the random device than ones that had been chosen directly. There is an obvious sense in which the random devices in these experiments are procedurally fair, and in which Blount's and Bolton et al.'s results are evidence of the relative acceptability of inequalities that are generated in this way. But if one thinks of a procedure as the 'rules of the game' for an interaction between individuals, these random devices are not games in the normal sense of the word. (Formally, they are degenerate games in which there are no decision nodes.) Given the decision to use a random device, the outcome of that device is not merely not intended by anyone; it is completely independent of any subsequent human decision.

Two early Ultimatum Game experiments reported by Hoffman and Spitzer (1985) and Hoffman et al. (1994) investigate attitudes to inequalities generated by procedures in which players compete for the right to be the first mover. The competition is either a game of Nim, players taking turns to move first (in the 1985 paper) or a general knowledge quiz (in the 1994 paper). These experiments find evidence that first movers offer less, and second movers are willing to accept less, when the role of first mover has been chosen by competition than when it has been selected at random. These are comparisons between different forms of procedural fairness. In contrast, our experiment compares fair procedures with procedures that are biased towards one participant and against another. Our experiment also differs from all its predecessors by using a new method to elicit attitudes to procedurally-generated inequalities.

3 Formalising the concept of procedural fairness

The intuitive idea of a fair procedure can be represented by a game (defined in terms of material consequences rather than utilities) in which all players have the same strategic opportunities. To keep things simple, we focus on the case of a finite two-player game in normal form. Initially, we restrict attention to games in which both players have the same number of strategies. Formally, each player $i \in \{1, 2\}$ has a set of m pure strategies $S_i = \{s_{i1}, \ldots, s_{im}\}$. For each player i there is a *payoff function* π_i which assigns a material payoff $\pi_i(s_{1k}, s_{2l})$ to each profile of pure strategies. We will normally interpret payoffs as quantities of some good – for example, money – that both players value positively. When we consider games that involve random processes or 'moves of nature', a payoff will be interpreted as a probability distribution over quantities of a positively-valued material good.

We propose that a procedure should be deemed to be fair if (but not only if), for all k, $l \in \{1, \ldots, m\}$, $\pi_1(s_{1k}, s_{2l}) = \pi_2(s_{1l}, s_{2k})$. In other words, player 1's payoff if he chooses his strategy k and player 2 chooses her strategy l is the same as player 2's payoff if she chooses her strategy k and player 1 chooses his strategy l. Notice that this condition uses the labelling convention that if strategies for the two players have the same number, they are 'the same'. To avoid this arbitrariness, we define a game as *procedurally fair* if it is possible to renumber strategies (separately for each player) in such a way that, after renumbering, $\pi_1(s_{1k}, s_{2l}) = \pi_2(s_{1l}, s_{2k})$ holds for all k and l. For example, consider the following Battle of the Sexes game:

Game 1:

		Betty	
		boxing	*opera*
Arthur	*boxing*	4, 3	0, 0
	opera	0, 0	3, 4

If we denote Arthur as player 1 and Betty as player 2, and if for each player we denote *boxing* by strategy 1 and *opera* by strategy 2, we find $\pi_1(s_{11}, s_{21}) \neq \pi_2(s_{11}, s_{21})$ and $\pi_1(s_{12}, s_{22}) \neq \pi_2(s_{12}, s_{22})$, contrary to the sufficient condition proposed above. But the same game can be re-described as one in which each player chooses between his or her *more preferred* and *less preferred* entertainments. If we keep the previous numbering of Arthur's strategies, but renumber Betty's so that *opera* becomes her strategy 1 and *boxing* becomes her strategy 2, $\pi_1(s_{1k}, s_{2l}) = \pi_2(s_{1l}, s_{2k})$ holds for all k and l: the game is procedurally fair.

Procedural fairness is a property of the rules of a game, considered without reference to players' actual behaviour in that game or to how the game would be played by rational individuals. For many types of interaction, it is possible to recognise the presence or absence of procedural fairness without even enumerating all the strategies available to the players. For example, consider a game of chess that is preceded by a coin toss to determine who plays as White; there is a money prize for the winner (shared equally in the event of a draw). Although the problem of computing optimal strategies for chess has so far defeated the most powerful computers, it is immediately obvious that this game is procedurally fair.

We now propose a concept of procedural bias that is similarly independent of considerations about rational play. The underlying intuition can be expressed by considering how one might induce a bias in favour of a particular player by changing the specification of an initially fair game. One method might be to increase one payoff for the player to be favoured while keeping other payoffs constant. For example, Game 2 has been constructed by taking a fair Battle of the Sexes game (i.e. Game 1, but without its specific labels) and substituting $\pi_1(s_{11}, s_{21}) = 5$ for $\pi_1(s_{11}, s_{21}) = 4$, thus favouring player 1. Another method might be to give the favoured player an additional, non-dominated strategy while keeping the rest of the game constant. For example, Game 3 has been constructed from Game 1 by giving player 1 such an additional strategy.

Game 2:

		player 2	
		strategy 1	strategy 2
player 1	strategy 1	5, 3	0, 0
	strategy 2	0, 0	3, 4

Game 3:

		player 2	
		strategy 1	strategy 2
	strategy 1	4, 3	0, 0
player 1	strategy 2	0, 0	3, 4
	strategy 3	2, 2	2, 2

Our formal definition of bias combines these two forms of 'favouring'. It is based on comparisons between two-player games that need not have the same number of strategies for each player. We will define bias towards player 1; bias towards player 2 can be defined symmetrically. In a typical game G (respectively G'), player 1 has m (m') strategies and player 2 has n (n') strategies. The players' payoff functions are π_1 (π_1') and π_2 (π_2'). For any two games G, G', we will say that G differs from G' by a *1-biased payoff revision* if $m = m'$ and $n = n'$ and if the only difference between the payoff functions for the two games is that for exactly one (k, l) pair, $\pi_1(s_{1k}, s_{2l}) > \pi_1'(s_{1k}, s_{2l})$.[1] We will say that G differs from G' by a *1-biased strategy expansion* if (i) $m = m' + 1$, (ii) $n = n'$, (iii) for all (k, l) pairs where $k < m$, $\pi_1(s_{1k}, s_{2l}) = \pi_1'(s_{1k}, s_{2l})$ and $\pi_2(s_{1k}, s_{2l}) = \pi_2'(s_{1k}, s_{2l})$, and (iv) there is no $k < m$ such that s_{1m} is weakly dominated by s_{1k}.[2] Intuitively, 1-biased payoff revisions and strategy expansions give player 1 weakly better opportunities to respond to any *given* strategy chosen by player 2.

Using these concepts, we define a game G as *procedurally biased towards* player 1 if there exists a procedurally fair game G' such that G can be constructed from G' by some sequence of 1-biased payoff revisions and 1-biased strategy expansions. Our definitions of fairness and bias are not exhaustive: many games that are not procedurally fair cannot be classified as biased in favour of either player. However, the classifications 'procedurally fair', 'procedurally biased towards player 1' and 'procedurally biased towards player 2' are mutually exclusive.[3] In general, the fact that a game is procedurally biased towards player 1 does not imply that, given Nash equilibrium play, player 1's payoff will be at least as great as

---

[1] For ease of exposition, we formulate our definitions for games in which there are no moves of nature. These definitions can be generalised by interpreting payoffs as probability distributions over material payoffs and interpreting strict (weak) inequalities as relations of strict (weak) first-order stochastic dominance.

[2] Strategy s_{1m} is weakly dominated by s_{1k} if, for all l, $\pi_1(s_{1k}, s_{2l}) \geq \pi_1'(s_{1m}, s_{2l})$, with a strict inequality for some l.

[3] Consider any game G that is biased towards player 1. Our definitions imply that either $m > n$, in which case G can be neither fair nor biased towards player 2, or $m = n$. Suppose the latter. Then there must be some procedurally fair game G' from which G can be constructed by a sequence of 1-biased payoff revisions. For any game with $m = n$, define *payoff advantage* (to player 1) as the sum of all possible payoffs to player 1 minus the sum of all possible payoffs to player 2. (In a game with moves of nature, 'sum' must be replaced by 'expected value when all strategy profiles are equally probable'.) In any procedurally fair game, payoff advantage is zero. Since any 1-biased payoff revision increases payoff advantage, payoff advantage in G must be strictly positive. Thus G cannot also be fair or biased towards player 2.

player 2's.[4] (Having more opportunities in a game can be a strategic disadvantage.) However, this implication holds for all zero-sum games.[5]

Procedural fairness can be interpreted as a very general principle of ex ante equality of opportunity. Many existing equal-opportunity principles, such as those proposed by Dworkin (1981) and Roemer (1998), imply procedural fairness in suitably-specified games. So too do the principles of 'liberal egalitarianism' (Cappelen et al, 2007) and 'choice egalitarianism' (Cappelen et al., 2013). However, our concept of procedural fairness is also compatible with Hayek's picture of a spontaneous order in which the idea of trying to match rewards to desert is out of place.[6]

It might be objected that, in any procedurally fair game, individuals are rewarded according to their talent and effort in playing that game. By definition, a procedurally fair game that did *not* present players with meaningful decisions between strategies would be a game of pure chance. If individuals understood that they were playing such a game, they would know that each of them began the game with the same probability distribution over final payoffs; in this ex ante sense, there would be no inequality. But (one might argue) if a procedurally fair game *does* involve meaningful strategic decisions, it must be possible to define a concept of decision-making skill that the game rewards; and this is a skill that plausibly depends on some combination of talent and cognitive effort. The implication is that, leaving aside the special case of purely random rewards, one cannot investigate attitudes to procedural fairness by using procedures that do not reward *any* form of skill or effort.

Nevertheless, there is a genuine difference between perceiving a given inequality to be fair because it rewards strategic skill and perceiving it to be fair because it is the outcome of a fair procedure. For example, one might take the first attitude to the inequality in prize

---

[4] For example, consider the game formed by substituting $\pi_1(s_{12}, s_{21}) = 4$ for $\pi_1(s_{12}, s_{21}) = 0$ in Game 1. This is a 1-biased payoff revision of a procedurally fair game, but its unique Nash equilibrium is (2, 3).

[5] Consider any zero-sum $m \times n$ game G that is procedurally biased towards player 1. By definition, $m \geq n$, and there exists a procedurally fair $n \times n$ game G' from which G can be constructed by a sequence of 1-biased payoff revisions and strategy expansions. Without loss of generality, number the strategies in the two games so that, in this sequence, a payoff revision to a strategy does not change its number. Consider any $k \leq n$. Since G' is procedurally fair, $\pi_1'(s_{1k}, s_{2k}) = \pi_2'(s_{1k}, s_{2k})$. By construction, $\pi_1(s_{1k}, s_{2k}) \geq \pi_2(s_{1k}, s_{2k})$; since G is a zero-sum game, this implies $\pi_1(s_{1k}, s_{2k}) \geq 0$. Thus, for every strategy available to player 2, player 1 has a response that gives a non-negative payoff. So, in every Nash equilibrium of G, player 1's payoff is non-negative and (because G is zero-sum) player 2's is non-positive.

[6] Compare Sugden's (2004) discussion of 'ex ante equality of opportunity' in a Hayekian economy.

money between the winner and the runner-up in a chess tournament while treating inequalities that result from differences between individuals' market decisions as underserved but morally acceptable on grounds of procedural fairness. The games we use in our experiment were chosen with the aim of making the dimension of procedural fairness as salient as possible relative to the dimensions of talent and effort.

4 Experimental design

Our experimental design elicited individuals' attitudes to inequalities that had been generated by different procedures. Part 1 of the experiment generated the inequality; Part 2 elicited attitudes towards it.

Our main comparison is between two treatments. In Part 1 of each of these treatments, participants were paired to compete in a series of identical zero-sum games. In the Fair Rule treatment, the games were procedurally fair; in the Unfair Rule treatment, they were procedurally biased towards one of the players. To allow a fuller comparison with the existing experimental literature, we also used a Real Effort treatment in which inequality was generated as the result of a procedure in which paired participants competed in a real effort task. Every competition ended with a winner and a loser; the winner received nine tickets for a specified lottery and the loser received three**.** Thus, all three treatments induced the same inequality of outcome between the two participants.

In Part 2 of each treatment, the paired participants played a vendetta game, adapted from the design introduced by Bolle et al. (2014). In this game, participants were given opportunities to take lottery tickets from their co-players, in alternating turns, to increase their holdings of tickets by a fraction of what they took. By using or not using these opportunities to take tickets, participants were able to reveal their attitudes to the inequality generated in Part 1. At the end of the experiment, for each pair of participants, one of the twelve lottery tickets was drawn at random. If that ticket had not been wasted in the vendetta game, its holder won a money prize.

We now describe the components of the experiment in more detail. The full instructions to participants can be found in the Online Appendix.

4.1 Assignment of roles

Each experimental session was assigned to one of the three treatments. At the beginning of each session, participants were admitted to the lab one by one and were asked to sit at any vacant computer terminal. At this stage, they were given no indication that the choice of where to sit might be significant. When the experiment began, participants were told that those sitting at odd-numbered seats would be ‘participant As’ and those at even-numbered seats would be ‘participant Bs’; each participant A would be randomly and anonymously matched with a participant B for the duration of the experiment. By allocating seats before the start of the experiment and in this unstructured way, we intended that the assignment of A and B roles would be separated in participants’ minds from the rules of the game or competition they faced in the experiment itself.

4.2 Part 1: the fair and unfair card games

In the Fair Rule and Unfair Rule treatments, each pair of participants played a series of *card games*, simulated on their computer screens. At the start of each game, each player was ‘dealt’ a card. Each card had a number of ‘points’, which could be any of the whole numbers from 1 to 100. Each of these numbers was equally likely at each deal (as if a new deck of 100 cards was used every time a card was dealt). Each player was offered a pre-specified number of opportunities to choose whether to stick with the card currently held or to return it and replace it with a newly-dealt card. In the Fair Rule treatment, participants A and B were both allowed to make up to three replacements in each game. In the Unfair Rule treatment, A was allowed to make no more than one replacement while B was allowed to make up to three. We will refer to A and B in this treatment as the *disadvantaged* and *advantaged* players respectively. (These terms were not used in the experiment itself.) During this stage of the game, neither player could see what cards her co-player was being dealt, or whether the co-player was using his replacement opportunities. After both players had chosen to stick with a card they had been dealt or had used up all their replacement opportunities, each saw the other’s final card and was shown how many replacement opportunities that player had used. The player holding the higher-numbered card was the winner of the game. If both cards had the same number, the game was a draw. The first player to win four games was the overall winner of the series of games: draws were not counted. (In fact, only 1.2 per cent of games were drawn.) A screen shot of the card game is shown in the Online Appendix.

Given our aim of investigating attitude to procedural fairness and unfairness, these games have a number of desirable features. Their rules are very easy to understand. The procedural fairness of the Fair Rule game, and the procedural bias of the Unfair Rule game (and the fact that this bias works to the advantage of the player with more replacement opportunities) are transparent and salient.[7] It is also obvious that each player's chance of winning depends partly on how she uses her replacement opportunities, and that in broad-brush terms, what a player needs to do is to replace relatively low-valued cards and to stick at relatively high-valued ones. Thus, the game induces purposeful engagement by the players. When one player learns that her co-player is using replacement opportunities, she can reasonably infer that he is trying to increase the probability that that he wins and she loses.[8] Nevertheless, the outcome of any single game, even under the unfair rules, is predominantly a matter of luck: in the Nash equilibrium of the unfair game, the probability of a win for the advantaged player is approximately 0.63, implying that this player's probability of winning the series is 0.77.[9] Although it would be very difficult for a participant to assess the exact balance of skill and luck in the game, the fact that luck plays an important part is easy to recognise. If, as one might expect, players' decisions are based on intuition rather than conscious calculation, playing the game involves little that could be construed as effort. Since, for each player, the series of games has only two possible outcomes ('win' or 'lose'), attitudes to risk are irrelevant, at least for rational players. For the reasons explained in Section 3, our games must provide *some* opportunity for skill or effort if they are to be useful in an investigation of procedural fairness. But we think it reasonable to assume that, under both the fair and unfair rules, there are no major differences of ability or personality traits between winners and losers.

4.3 Part 1: the real effort task

Real effort tasks, in which experimental subjects are rewarded for their performance in burdensome low-skill tasks, are commonly used to investigate desert-based concepts of fairness (e.g. Burrows and Loomes, 1994; Fahr and Irlenbusch, 2000; Konow, 2000). Part 1 of our Real Effort treatment was designed to create inequalities that could be construed as rewarding differences in effort.

---

[7] That the Unfair Rule game is procedurally biased towards player B can be shown formally by considering any given strategy s_A for A. There is an exactly equivalent strategy for B: replace the first two cards, irrespective of their values, and then use s_A at the final replacement opportunity. But B also has many undominated strategies that are not available to A.

[8] Only three of the 104 participants in the Fair Rule treatment and only four of the 150 participants in the Unfair Rule treatment failed to use any replacement opportunities.

[9] These numbers are approximations because they are calculated for a game in which card numbers are uniformly distributed over the real interval [0, 100]. One of our reasons for using a series of games rather than a single one was that we wanted the game itself to be simple, while giving the advantaged player a high probability of winning overall.

We used the *encryption task*, as developed by Erkal et al. (2011). For each pair of matched participants, each member of that pair was given an encryption table which assigned a number to each letter of the alphabet in a random order. She was then presented with words in a predetermined sequence and was asked to encrypt them by substituting the letters with numbers using the encryption table. All participants were given the same words to encode in the same order. After a participant encoded a word, the computer would tell her whether the word had been encoded correctly. If it had been encoded wrongly, she would be asked to check her codes and correct them. Each time she encoded a word correctly, she was given another word to encode. This process continued for six minutes. The participant who had encoded more words in this period was declared the winner. If there was a tie, the winner was the participant who had encoded the words in a shorter time. A screen shot of the real effort task is shown in the Online Appendix.

4.4 Part 2: the vendetta game

Part 2 took the same form in all three treatments. Participants remained in the same pairs as in Part 1, but now were referred to as 'the winner in Part 1' and 'the loser in Part 1'. The game involved twelve lottery tickets, numbered from 1 to 12. Initially, nine of these tickets, selected at random, were assigned to the Part 1 winner; the remaining three were assigned to the loser. The players moved in turn, the loser moving first. When it was a player's turn to move, she was asked to choose whether she wanted to take lottery tickets from her co-player, and if so, how many tickets to take. Tickets had to be taken in blocks of three (so the number of tickets taken could be three, six or nine, up to as many as the co-participant held at the time). For each block of three tickets that the player took, one of those tickets was transferred to her and the other two were wasted (i.e. lost to both players). Players always had the option of not taking any tickets. The game ended if one of two cases applied. The first case occurred if one or both players were still able to take tickets, but the player(s) who was (were) able to do this had chosen not to do so on two consecutive moves. The second case occurred if both players held less than three tickets, and therefore no positive multiple of three tickets could be taken from either of them. Figure 1 shows all the distributions of tickets that can be reached in the game. The starting point is (3, 9), i.e. the loser holds three tickets and the winner holds nine. For each possible distribution of tickets, the solid arrows show the possible moves by the loser from that point in the game; any continuous sequence of these arrows is a possible move. For example, if the current distribution is (3, 9) and the loser has an opportunity to take, she can move to (4, 6), (5, 3) or (6, 0). Similarly, the broken arrows show possible moves by the winner. The only distribution at which no further taking moves are possible is (0, 2).

[Figure 1 near here]

This game has the same general structure as the vendetta games studied by Bolle et al. (2014), but it was framed in more concrete terms and displayed in a more intuitive visual form. In Bolle et al.'s experiment, states of the game were described as numerical probabilities of winning, which could be changed according to algebraic rules. In our design, these states were represented as distributions of physical objects between three 'baskets' – one for each player, and a 'bin' for wasted tickets. Players' decisions transferred specific objects between these baskets. Figure 2 shows a screen shot of the beginning of the game.

[Figure 2 near here]

If players are rational and self-interested and if this is common knowledge, the distribution (4, 6) is the unique backward-induction solution. (The first stage of the induction shows that, if (6, 0) or (3, 1) is reached, the winner will move the game to (0, 2); hence, no move will be made from (2, 4). The second stage shows that, if (5, 3), (0, 10) or (1, 7) is reached, one of the players will move the game to (2, 4); hence, no move will be made from (4, 6). The third stage shows that, starting from (3, 9), the loser will move the game to (4, 6).) Other outcomes are possible for rational players who act on social preferences. For example, if both players are mildly altruistic (specifically, if each player treats one unit of the other's payoff as equivalent to more than one third of a unit of her own), the game will stay at (3, 9). But we must emphasise that our design does not presuppose any rationality assumptions, and is not intended to investigate alternative explanations of taking behaviour. We designed the experiment in the light of Bolle et al.'s finding that vendetta games induce substantial frequencies of taking moves, even when those moves are contrary to received theories of rational play. For our purposes, the vendetta game is an instrument for measuring players' attitudes (whether rational or not) to the fairness or unfairness of different procedures. The relevant experimental control comes from the fact that the same vendetta game is used in conjunction with different procedures, all of which generate the same payoff inequality.

As far as we know, our experiment is the first to use vendetta games in this way. As we pointed out in Section 2, the usual method of eliciting attitudes to fairness is to ask individual respondents to choose between alternative payoff allocations. If attitudes to fairness could be represented as stable preferences over payoff distributions, the usual method would certainly be more controlled than a vendetta game; but the validity of that representation is not self-evident. In many real-world situations in which issues of fairness are salient, ordinary individuals do not have opportunities to make unilateral, uncontested changes to the distribution of income. The implications of this fact are now widely recognised in relation to acts of punishment. Fehr and Gächter (2000, 2002) show that the existence of one-sided punishment opportunities increases contribution levels in public good games. However, Nikiforakis (2008) shows that when both punishment and counter-punishment are allowed in a public good game, cooperators' willingness to punish decreases, leading to the breakdown of cooperation. In the presence of counter-punishment opportunities, people also reveal strong desires to reciprocate punishment (Cinyabuguma *et al.*, 2006; Denant-Boemont *et al.*, 2007). These findings cast doubt on the external validity of experimental designs that provide one-sided opportunities for punishment or redistribution. The real-world situations in which people are able to express resistance to procedural unfairness may be better modelled by vendetta games, and the attitudes that are expressed in these situations may correspond with 'hot' emotions of resentment and anger rather than stable distributional preferences. Because we wanted to pick up these hot emotions, we used a design in which vendetta games were played in real time, rather than using the strategy method to elicit players' conditional decisions at all nodes in the game tree.

4.5 Final earnings

At the end of each session of the experiment, twelve tickets numbered from 1 to 12 were put into a bag and a participant was asked to draw one at random. This draw determined the winning ticket number for every vendetta game in that session. In each pair, if the winning ticket was in the basket of one of the participants, that participant received a prize of £24; the other participant earned nothing from the game. If the winning ticket was in the bin, both participants earned nothing from the game. In addition, every participant was paid a participation fee of £3.

4.6 Implementation

The experiment was conducted between November 2015 and January 2016 at the CBESS Experimental Laboratory at the University of East Anglia. Participants were recruited from the general student population via the CBESS online recruitment system (Bock *et al.*, 2012). The experiment was programmed and conducted with the experimental software z-Tree

(Fischbacher, 2007). We ran 18 sessions in total: six for the Fair Rule treatment (with 104 participants), eight for the Unfair Rule treatment (150 participants), and four for the Real Effort treatment (72 participants). We needed more participants in the Unfair Rule treatment than the Fair Rule treatment because our main aim was to compare the vendetta game behaviour of Fair Rule participants with that of Unfair Rule participants from series of card games that had been won by the advantaged player; as explained in Section 4.2, we expected around 20 to 25 per cent of these series to be won by the disadvantaged player. We recruited fewer participants for the Real Effort treatment, as this was intended only to provide a point of comparison with the existing literature. Participants' ages ranged from 18 to 63; approximately 60 per cent were female.[10] The experiment lasted about 50 minutes. Every participant earned either £3 (the participation fee) or £27 (that fee plus the £24 lottery prize). Average earnings were £10.67 per participant.

5 Hypotheses

Informally stated, our fundamental hypothesis is that there is greater willingness to accept inequality if it results from a procedurally fair game than if it results from a game that was procedurally biased in favour of the player who ended up as the winner. Thus, we are primarily concerned with behaviour in the vendetta game played in Part 2 of the experiment, interpreted as revealing participants' attitudes to the inequality created in Part 1. That inequality could be created in four different ways – through a Fair Rule game (FR), through an Unfair Rule game that was won by the advantaged player (URA), through an Unfair Rule game that was won by the disadvantaged player (URD), or through Real Effort task (RE). It is convenient to refer to these as different 'treatments' applied to a single game.[11] Our fundamental hypothesis requires comparisons between the FR and URA treatments.

This hypothesis can be firmed up in various ways, depending on whether 'acceptance' is interpreted as an individual or collective phenomenon and on whether it is measured by players' propensities to stick at the initial token allocation or (inversely) by their propensities

[10] The gender question allowed 'prefer not to say' as an answer. Of the 326 participants, 123 identified as male and 196 as female; 7 preferred not to say.

[11] This is a slight misnomer, because the allocation of participants between URA and URD 'treatments' was partly the result of decisions made in Part 1. But, as explained in Section 4.2, we think it is a reasonable approximation to treat this allocation as if it were random.

to take from one another in the vendetta game as a whole. We test the following two hypotheses about collective behaviour:

> *Hypothesis 1 (collective sticking at the initial allocation):* The vendetta game is more likely to end at the (3, 9) allocation in the FR treatment than in the URA treatment.
>
> *Hypothesis 2 (collective taking):* In total, more tokens are wasted in the URA treatment than in the FR treatment.

Intuitively, one might expect non-acceptance of inequality to take different forms for winners and losers. A loser can perceive at least some acts of taking both as a protests against the initial inequality and as partial rectifications of it. Thus, if inequality is perceived as less acceptable in the URA treatment than in the FR treatment, we should expect losers' willingness to take to be greater in URA games. Our prior expectation was that there would be an opposite effect for winners, that is, a tendency for them to view modest amounts of taking by losers as more legitimate, and so less deserving of retaliation, in URA games. To the extent that winners engage in the apparently myopic act of taking at (3, 9) – adding to the initial inequality by moving the game to a position where their opponents have nothing to lose by counter-taking – one might expect that behaviour to be less likely, the less acceptable the initial inequality. These expectations are encapsulated in the following hypotheses about individual behaviour:

> *Hypothesis 3 (individual sticking at the initial allocation)*: (a) For losers, sticking at the (3, 9) allocation is less likely in the URA treatment than in the FR treatment; and (b) for winners, sticking at that allocation is more likely in the URA treatment than in the FR treatment (unless there is 100 per cent sticking in both).
>
> *Hypothesis 4 (individual taking)*: (a) For losers, the overall propensity to take is greater in the URA treatment than in the FR treatment; and (b) for winners, the overall propensity to take is lower in the URA treatment than in the FR treatment.

The concept of 'overall propensity to take' in Hypothesis 4 needs some explanation. Because of the interactive nature of vendetta games, and because we can observe players' behaviour only at the nodes in the game they actually reach, disentangling winners' and losers' underlying attitudes to taking is not straightforward. The fundamental problem is that which nodes are reached by one player depends in part on the behaviour of the other. If we are to identify determinants of behaviour separately for winners and losers, we need to control for this effect. In principle, it is possible to test for overall cross-treatment differences in taking propensities by carrying out a separate test at each decision node.[12] Consider any given decision node *N* for the winner (for example, the winner's second opportunity to take at (4, 6), reached after the loser took three tokens at her first opportunity at (3, 9)). Suppose we observe behaviour at this node in both FR and URA games, and we want to test whether this observation is consistent with the null hypothesis that the distribution of winners' behaviour strategies (i.e. strategies that specify a player's behaviour at *every* node in the game) is the same for FR and URA players. A winner can reach *N* only if a particular combination of previous moves is played. But, given the null hypothesis and provided that (within FR and URA games considered separately) there is no correlation between the strategies of matched players, the distribution of winners' strategies conditional on reaching *N* is the same for FR and URA games. Thus, if behaviour at *N* is significantly different in the two games, that is evidence against the null hypothesis. Intuitively, the implication is that Hypothesis 4 should be read as referring to overall patterns in the results of the node-specific tests we have described. In Section 6.3 we will explain how we used this idea to design practicable tests of that hypothesis.

Our experiment also allows tests, analogous with those of Hypotheses 1 to 4, of differences between the FR and RE treatments. We do not propose any formal hypotheses about differences between these treatments. Nevertheless, comparisons between these treatments are useful for calibrating the acceptability of inequalities generated by procedurally fair games relative to that of inequalities generated by a standard experimental procedure that is usually interpreted as rewarding individuals according to a principle of desert.

6 Results

For completeness, we report data for all four treatments. Our primary concern is with Hypotheses 1 to 4, which involve comparisons between the FR and URA treatments. We also report tests of differences between the FR and RE treatments. As explained in Section

[12] Following the standard practice in game theory, we define nodes in relation to game trees. Since the vendetta game is a game of perfect information, each decision node for a specified player has a unique history of previous moves by all players.

4.6, the URD games exist only as a by-product of the procedure by which the URA games were created; the number of these games (17) is too small for useful analysis.

6.1 Vendetta game outcomes

Table 1 summarizes the outcomes of the vendetta game in the four treatments. To get a broad-brush picture of how the game was played, we begin by looking at the aggregate data in the final column of the table.

[Table 1 near here]

Recall that, if players act on rational self-interest and if this is common knowledge, the outcome of the game's unique backward-induction solution is the (4, 6) allocation. In fact, this was the end point of only 12.9 per cent of games. In 32.5 per cent of games, the game ended at the initial (3, 9) allocation. In these games, losers chose not to take, in preference to the move (taking three tokens) that is specified by the backward-induction solution. This behaviour can be rationalized in various ways – for example, in terms of altruism on the part of the loser, respect for entitlements ('taking' might be viewed as a kind of stealing), a perception of not taking as the default option, or a belief that the winner might retaliate against a loser who took three tokens.[13] But notice that 47.9 per cent of games ended with allocations that were strictly worse for both players than both (3, 9) and (4, 6). In these cases, the behaviour induced by the vendetta game seems dysfunctional, both for each player individually and for the two players collectively. Our experiment was not designed to investigate players' motivations, but we conjecture that this mutually damaging behaviour was driven by hot emotions of resentment and anger (and perhaps, in the case of winners who took at (3, 9), greed), and by each player's failure to anticipate the full effects of the other player's experiences of those emotions.[14] This evidence suggests that the responses to inequality that are picked up by the vendetta game are fundamentally different from those that individuals report when they take on the role of a social planner.

---

[13] In the context of the experiment, a belief of this kind would in fact have been justified. Of the 49 winners who reached (4, 6), 14 chose to take at the first opportunity; of the 22 winners who, having declined the first opportunity, were given a second, 5 chose to take.

[14] The relatively large proportion of games ending at (2, 4) may be evidence of some minimally sophisticated reasoning by players who had experienced the effects of taking and counter-taking. If a loser takes at (2, 4), the winner has nothing to lose by retaliation, and the loser has no way of counter-retaliating.

We now consider cross-treatment comparisons. In the FR treatment, 44.2 per cent of pairs stuck at the initial (3, 9) allocation, compared with only 22.4 per cent in the URA treatment; this difference is significant in a two-tailed chi-squared test (χ^2= 5.928, p = 0.015).[15] Hence:

> *Result 1*: Consistently with Hypothesis 1, the vendetta game was more likely to end at the (3, 9) allocation in the FR treatment than in the URA treatment.

In the RE treatment, 38.9 per cent of pairs stuck at (3, 9), which is not significantly different from the FR treatment (χ^2= 0.249, p = 0.618).

Table 1 shows the average number of tickets held at the end of the game by each player in each treatment. For both winners and losers, these numbers are markedly greater in the FR treatment (5.87 and 2.40 respectively) than in the URA treatment (4.67 and 1.47); the RE treatment (5.58 and 2.36) is similar to FR. Figure 3 shows the cumulative distributions of final total holdings (i.e. the sum of the holdings of the two players) in the three treatments. Final total holdings in the URA and FR treatments are significantly different from one another (Mann-Whitney p = 0.046; the probability that a randomly selected URA subject has smaller final holdings than a randomly selected FR subject is 0.607). Hence:

> *Result 2*: Consistently with Hypothesis 2, more tokens were wasted in the URA treatment than in the FR treatment.

There is no significant difference in total final holdings between the RE and FR treatments (Mann-Whitney p = 0.679; the probability that a randomly selected RE subject has smaller final holdings than a randomly selected FR subject is 0.525).

[Figure 3 near here]

6.2 Sticking at the initial allocation by winners and losers

Table 2 gives a breakdown of winners' and losers' decisions about whether to take at the initial allocation. A player's behaviour is classified according to the last move she made at (3, 9). This move is $take_1$ if she took (any positive number of) tokens at her first opportunity. It is $pass_1$ if she passed (i.e. took no tokens) at her first opportunity, and the other player then

[15] Throughout the paper, significance levels are reported for two-tailed tests.

immediately took tokens.[16] It is $take_2$ if she passed at her first opportunity, the other player then passed, and she then took at her second opportunity. It is $pass_2$ if she passed at her first opportunity, the other player then passed, and she then passed at her second opportunity. Notice that if a player's last move is $pass_1$, we do not know how she would have used a second opportunity to take. But a player whose last move is $take_1$ or $take_2$ has unambiguously chosen to take, and one whose last move is $pass_2$ has unambiguously chosen to pass. We therefore define the *sticking rate* for a given role (loser or winner) as the number of $pass_2$ moves as a proportion of the total of $take_1$, $take_2$ and $pass_2$; n is the total number of moves of all four types (i.e. the number of players in the relevant role who had at least one opportunity to take).

[Table 2 near here]

For losers, the sticking rate is higher in the FR treatment (47.1 per cent) than in the URA treatment (30.4 per cent); this difference is significant at the 10 per cent level (χ^2= 3.150, p = 0.076). But contrary to our prior expectation, we find a significant difference in the same direction for winners.[17] The sticking rate is 92.0 per cent for FR winners and 68.4 per cent for URA winners (χ^2= 4.035, p = 0.045). Hence:

> *Result 3*: Consistently with Hypothesis 3, we find weak evidence that losers were more likely to stick at (3, 9) in the FR treatment than in the URA treatment. But, contrary to that hypothesis, winners were more likely to stick at (3, 9) in the FR treatment.

Comparisons of sticking rates between FR and RE treatments show no significant difference either for losers (χ^2= 0.573, p = 0.450) or for winners (χ^2= 1.181, p = 0.277).

6.3 Winners' and losers' taking propensities

As explained in Section 5, an ideal test for cross-treatment differences in taking propensities would be based on separate comparisons at each decision node. However, since the vendetta game has 82 decision nodes for winners and 46 for losers, disaggregating behaviour to the decision-node level would produce numbers that are much too small for useful analysis. Our

[16] If the other player was the winner, this was that player's first opportunity to take; if he was the loser, it was his second.

[17] Sticking rates for winners should be interpreted with some caution, because of the small numbers of winners who chose to take at (3, 9). Intuitively, a winner who takes at (3, 9) is making a move that is both unprovoked and rash: it leaves the loser with no tokens, and with opportunities to retaliate.

analysis is therefore based on a coarser disaggregation of the data. The results of this analysis represent our best attempt to disentangle the behaviour of losers and winners.

We define a *taking opportunity* for a given player as a set of decision nodes for that player, specified by the current allocation of tokens and (if the allocation is one at which it is possible for both loser and winner to take tokens) whether the relevant player is facing her first or second opportunity to take. (At an allocation such as (0, 6), at which only one player is able to take, the distinction between first and second opportunities has no strategic significance.) Using L and W subscripts to refer to loser and winner, and 1 and 2 subscripts to refer to first and second opportunities, the set of taking opportunities for the loser is O_L = $\{(3, 9)_{L1}, (3, 9)_{L2}, (4, 6)_{L1}, (4, 6)_{L2}, (5, 3)_{L1}, (5, 3)_{L2}, (0, 10)_L, (1, 7)_L, (2, 4)_L\}$; the corresponding set for the winner is $O_W = \{(3, 9)_{W1}, (3, 9)_{W2}, (4, 6)_{W1}, (4, 6)_{W2}, (5, 3)_{W1}, (5, 3)_{W2}, (6, 0)_W, (3, 1)_W\}$. In defining taking opportunities in this way, we ignore the history of how a given allocation was reached. If (as is the case for (1, 7), (2, 4) and (3, 1)) a given allocation can be reached by an immediately preceding taking move by either player, we implicitly ignore whether, at that allocation, the loser or the winner has the first opportunity to take. A full breakdown of behaviour at each taking opportunity is reported in Tables A1–A4 in the Appendix.

Using data from these tables we construct an index of 'excess taking' for each player in the FR and URA treatments. Intuitively, this index measures each player's taking propensity in the game as a whole, relative to the average taking propensity of players in the same role (loser or winner) in those treatments, taken together. We now explain how this index is defined for losers; the definition for winners follows the same principles.

Consider any taking opportunity k in O_L, and any individual i who played as a loser in the FR or URA treatment. Let a_k be the observed average number of tokens taken by players who faced that taking opportunity in the FR and URA treatments, taken together. We define a variable t_{ik} as follows. If i faced opportunity k, t_{ik} is the number of tokens taken by i at k; otherwise, $t_{ik} = a_k$. The null for a test of part (a) of Hypothesis 4 is that the behaviour strategies of losers in the two treatments are drawn from the same distribution. Under this null hypothesis, the expected value of $t_{ik} - a_k$ for any given k is zero in both treatments, irrespective of whether winners' behaviour differs between those treatments. Suppose that, for each taking opportunity in O_L, we fix any strictly positive and finite weight q_k, and then define $X_i = \sum_{k\in O_L}(t_{ik} - a_k)q_k$. Summing over all individuals in both treatments, and whatever weights are used, $\Sigma_i(t_{ik} - a_k) \equiv 0$ for each opportunity k, and hence X_i necessarily has a mean of zero for the two treatments taken together. Under the null hypothesis, the expected value of X_i is zero in each treatment taken separately.

As a convention, we define the *index of excess taking* for each loser *i* as the value of X_i if, for each taking opportunity *k*, q_k is the proportion of FR and URA games in which that opportunity *k* was faced. Intuitively, this index weights each taking opportunity by the probability that it is reached in an 'average' game. The value of this index for any given player *i* can then be thought of as an estimate of the difference between *i*'s taking propensity and the taking propensity of an 'average' player, based only on observations of *i*'s behaviour. We interpret part (a) of Hypothesis 4 as predicting that indices of excess taking for losers are higher in the URA treatment than in the FR treatment. Part (b) is interpreted as predicting that indices of excess taking for winners are lower in the URA treatment.

Our tests of these hypotheses are reported in the first column of Table 3. The entries in the table report, separately for losers and winners, the difference between the mean value of the index of excess taking in the URA and FR treatments; positive values indicate a greater taking propensity in the URA treatment. The test statistic is the *p*-value for a *t*-test with bootstrap. Recall that, summing over the behaviour of winners and losers, total taking in the URA treatment (5.11 tokens per game) was significantly greater than in the FR treatment (3.73 tokens per game). Roughly speaking, a comparison between the 'loser' and 'winner' entries in the first column of Table 3 is informative about the relative contributions of losers and winners to this overall difference in taking behaviour. For losers, as one would expect, the propensity to take is greater in the URA treatment than in the FR treatment, but the difference is not statistically significant. More surprisingly, the propensity to take *by winners* is also greater in the URA treatment, and this difference is significant at the 10 per cent level. Hence:

> *Result 4:* Our index-based test finds no significant support for Hypothesis 4(a), i.e. the prediction that losers' taking propensities are greater in the URA treatment than in the FR treatment. We find weak evidence that, contrary to Hypothesis 4(b), winners' taking propensities are greater in the URA treatment.

The second column of Table 3 reports similar comparisons between propensities to take in the RE and FR treatments. There are no significant differences between treatments, either for losers or winners.

[Table 3 near here]

7 Discussion

The main objective of our experiment was to investigate individuals' attitudes to inequalities of outcome that have not been directly chosen by anyone, but instead result from rule-governed interactions in which those individuals have participated and in which each of them has acted on self-interest. Our experiment was designed to elicit 'hot' emotions of resentment and anger about inequality rather than 'cool' social preferences or distributional judgements. Our main finding (expressed in Results 1 and 2) is that inequalities are perceived as more acceptable if the interaction that generates them is procedurally fair than if it is biased. None of our measures of acceptability revealed any significant difference between the FR and RE treatments. In other words, inequalities generated by self-interested behaviour in a fair competition whose outcomes were largely determined by chance were perceived in much the same way as inequalities that could be interpreted as rewards for differences in effort. We interpret this as evidence that people are relatively tolerant of inequalities that no one has consciously chosen, but which result from interactions that are governed by procedurally fair rules.

Our initial expectation was that, in the URA treatment, the driving force for wealth-destruction in the vendetta game would be resentment and anger *felt by the losers of the card game* – that is, by players who had been disadvantaged by the unfair rules under which that game had been played. We expected an opposite effect for winners in that treatment. Our intuition was that winners would realise that they had already benefited from unfair rules, and so would be less likely to initiate taking than in the FR treatment, and less likely to retaliate against taking moves by their co-players. Surprisingly, however, we found no evidence that advantaged winners were inhibited about taking. We recognise that our findings about the separate behaviour of losers and winners are less firm than those about their combined behaviour, both because of the smaller sample sizes and because of the difficulty of disentangling the behaviour of players who interact in a sequential game. But Results 3 and 4 both suggest that winners were *more* likely to take in the URA treatment than in the FR treatment.

As this effect was unexpected, we can offer only post hoc conjectures about how it might be explained. Why would the fact that a person has already benefitted from unfairness make him or her more willing to act aggressively against the victim of that unfairness? There may be some analogy with psychology experiments which have famously shown that ordinary people can act with (what they have good reason to believe to be) cruelty towards others when that cruelty has been licensed by an apparently legitimate authority (Millgram, 1963; Zimbardo, 2007). In a less extreme form, our Unfair Rule treatment may have been perceived as licensing subjects to override their inhibitions against taking advantage of their co-participants. Or, putting this more charitably, interacting with others in procedures that are governed by fair rules may tend to prime social norms that impose fairness constraints on individual behaviour. By failing to activate such norms, the Unfair Rule treatment may have made it psychologically easier for both losers and winners to engage in behaviour that was collectively dysfunctional.

To repeat, we do not want to place too much weight on our findings about winners and losers considered separately. We believe that the main contribution of this paper is to highlight the importance of procedural fairness. Individuals have negative attitudes to procedural unfairness, independently of any social preferences for social welfare, equality of outcomes, or rewards for effort. Those attitudes may induce emotions of anger, resentment and greed that are revealed in behaviour that is socially costly.

Table 1: Outcomes of vendetta games

	Fair Rule (FR)	Unfair Rule with advantaged winner (URA)	Unfair Rule with disadvantaged winner (URD)	Real Effort (RE)	All
Allocation of tickets:					
(3, 9)	23 (44.2%)	13 (22.4%)	3 (17.7%)	14 (38.9%)	53 (32.5%)
(4, 6)	4 (7.7%)	10 (17.2%)	3 (17.7%)	4 (11.1%)	21 (12.9%)
(5, 3)	4 (7.7%)	1 (1.7%)	0 (0%)	3 (8.3%)	8 (4.9%)
(6, 0)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
(0, 10)	0 (0%)	0 (0%)	1 (5.9%)	0 (0%)	1 (0.6%)
(1, 7)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
(2, 4)	10 (19.2%)	16 (27.6%)	2 (11.8%)	6 (16.7%)	34 (20.9%)
(3, 1)	0 (0%)	2 (3.5%)	0 (0%)	0 (0%)	2 (1.2%)
(0, 2)	11 (21.2%)	16 (27.6%)	8 (47.6%)	9 (25.0%)	44 (27.0%)
Average number of tickets held by the winner	5.87	4.79	4.67	5.58	5.30
Average number of tickets held by the loser	2.40	2.10	1.47	2.36	2.19
Number of pairs	52	58	17	36	163

Table 2: Sticking at the initial allocation

Loser's last move at (3, 9)

	FR	URA	URD	RE
n	52	58	17	36
last move:				
$take_1$	24	35	12	17
$pass_1$	1	2	1	0
$take_2$	3	4	1	5
$pass_2$	24	17	3	14
sticking rate (%)	47.1	30.4	18.8	38.9

Winner's last move at (3, 9)

	FR	URA	URD	RE
n	28	23	5	19
last move:				
$take_1$	1	2	1	0
$pass_1$	3	4	1	5
$take_2$	1	4	0	0
$pass_2$	23	13	3	14
sticking rate (%)	92.0	68.4	75.0	100.0

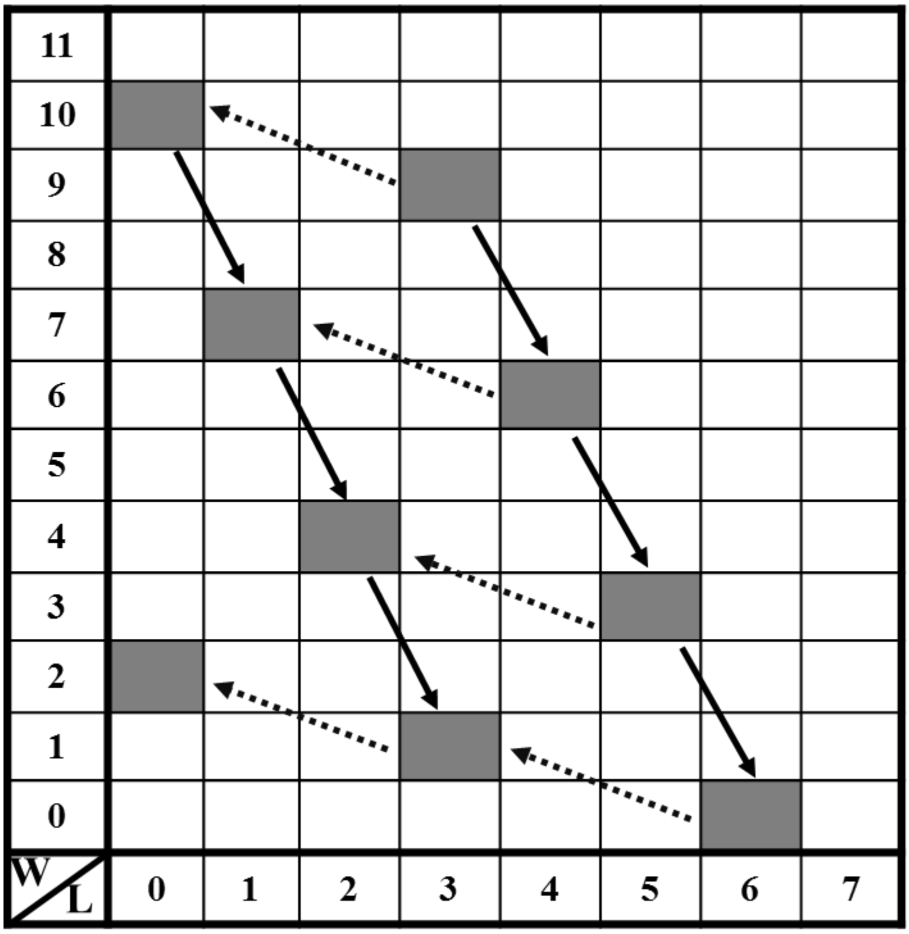
FR = Fair Rule treatment; URA = Unfair Rule treatment with advantaged winner; URD = Unfair Rule treatment with disadvantaged winner; RE = Real Effort Treatment. Sticking rate = $pass_2$ as percentage of ($take_1$ + $take_2$ + $pass_2$).

Table 3: Excess taking relative to Fair Rule treatment

	URA	RE
Loser	0.553 ($p = 0.354$)	–0.291 ($p = 0.671$)
Winner	0.147 ($p = 0.083$)	0.105 ($p = 0.457$)

URA = Unfair Rule treatment with advantaged winners; RE = Real Effort Treatment. This table shows the net excess of the mean index of excess taking for the URA (respectively RE) treatment over the mean index for the Fair Rule treatment; *p* values are from a *t*-test with bootstrap.

Figure 1: Schematic representation of vendetta game



Notes: The number of tickets held by the loser (winner) is shown on the horizontal (vertical) axis. The shaded cells are ticket distributions that can be reached in the game. Continuous sequences of solid (broken) arrows correspond with possible moves by the loser (winner).

Figure 2: Screen shot of vendetta game

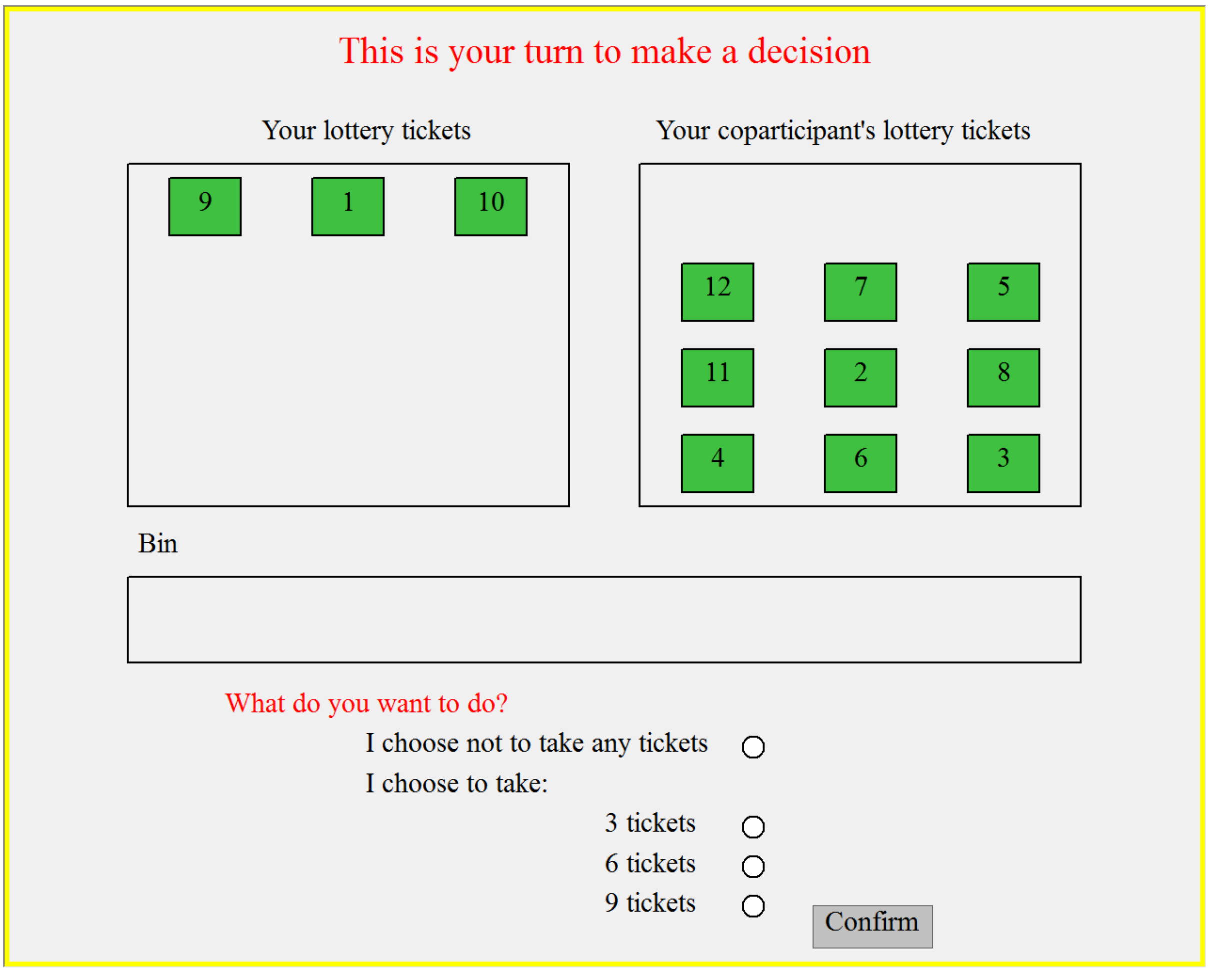
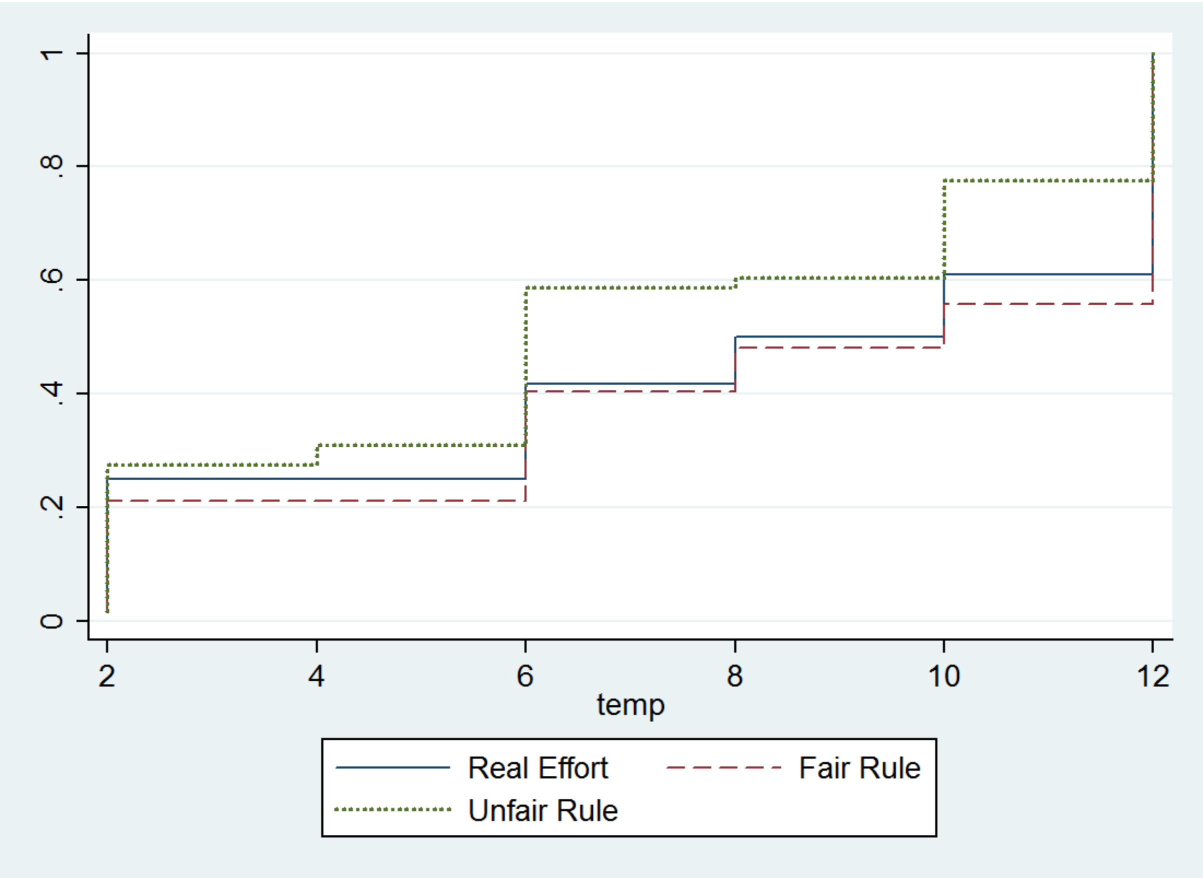


Figure 3: Cumulative distributions of final total holdings



Appendix: Behaviour at each taking opportunity

[Intended for print publication at end of main paper.]

These tables show, for each taking opportunity in each treatment, the number of players who faced that opportunity (*n*), and the distribution of these players' decisions between 'take 0', 'take 3', 'take 6' and 'take 9'.

Table A1: Fair Rule Treatment (FR)

Loser					
Taking opportunity	*n*	take 0	take 3	take 6	take 9
$(3,9)_{L1}$	52	28	12	7	5
$(3,9)_{L2}$	27	24	1	2	0
$(4,6)_{L1}$	10	8	2	0	
$(4,6)_{L2}$	6	4	2	0	
$(5,3)_{L1}$	5	5	0		
$(5,3)_{L2}$	4	4	0		
$(0,10)_{L}$	2	0	0	0	2
$(1,7)_{L}$	5	0	4	1	
$(2,4)_{L}$	13	0	3		
Winner					
Taking opportunity	*n*	take 0	take3	take 6	
$(3,9)_{W1}$	28	27	1		
$(3,9)_{W2}$	24	23	1		
$(4,6)_{W1}$	13	10	3		
$(4,6)_{W2}$	8	6	2		
$(5,3)_{W1}$	13	5	8		
$(5,3)_{W2}$	5	4	1		
$(6,0)_{W}$	5	0	1	4	
$(3,1)_{W}$	7	0	7		

Table A2: Unfair Rule treatment with advantaged winner (URA)

Loser					
Taking opportunity	*n*	take 0	take 3	take 6	take 9
$(3,9)_{L1}$	58	23	18	12	5
$(3,9)_{L2}$	21	17	3	1	0
$(4,6)_{L1}$	15	12	3	0	
$(4,6)_{L2}$	10	10	0	0	
$(5,3)_{L1}$	2	2	0		
$(5,3)_{L2}$	2	1	1		
$(0,10)_{L}$	6	0	0	2	4
$(1,7)_{L}$	8	0	5	3	
$(2,4)_{L}$	21	16	5		
Winner					
Taking opportunity	*n*	take 0	take 3	take 6	
$(3,9)_{W1}$	23	21	2		
$(3,9)_{W2}$	17	13	4		
$(4,6)_{W1}$	21	15	6		
$(4,6)_{W2}$	12	10	2		
$(5,3)_{W1}$	16	2	14		
$(5,3)_{W2}$	2	2	0		
$(6,0)_{W}$	6	0	1	5	
$(3,1)_{W}$	13	2	11		

Table A3: Unfair Rule treatment with disadvantaged winner (URD)

		Loser			
Taking opportunity	*n*	take 0	take 3	take 6	take 9
(3,9)$_{L1}$	17	5	5	3	4
(3,9)$_{L2}$	4	3	0	1	0
(4,6)$_{L1}$	3	3	0	0	
(4,6)$_{L2}$	3	3	0	0	
(5,3)$_{L1}$	0	0	0		
(5,3)$_{L2}$	0	0	0		
(0,10)$_{L}$	1	1	0	0	0
(1,7)$_{L}$	2	0	0	2	
(2,4)$_{L}$	4	2	2		
		Winner			
Taking opportunity	*n*	take 0	take 3	take 6	
(3,9)$_{W1}$	5	4	1		
(3,9)$_{W2}$	3	3	0		
(4,6)$_{W1}$	5	3	2		
(4,6)$_{W2}$	3	3	0		
(5,3)$_{W1}$	4	0	4		
(5,3)$_{W2}$	0	0	0		
(6,0)$_{W}$	4	0	0	4	
(3,1)$_{W}$	4	0	4		

Table A4: Real Effort Treatment (RE)

Loser					
Taking opportunity	*n*	take 0	take 3	take 6	take 9
$(3,9)_{L1}$	36	19	6	9	2
$(3,9)_{L2}$	19	14	4	1	0
$(4,6)_{L1}$	7	5	2	0	
$(4,6)_{L2}$	4	4	0	0	
$(5,3)_{L1}$	5	3	2		
$(5,3)_{L2}$	3	3	0		
$(0,10)_{L}$	0	0	0	0	0
$(1,7)_{L}$	4	0	3	1	
$(2,4)_{L}$	10	6	4		
Winner					
Taking opportunity	*n*	take 0	take 3	take 6	
$(3,9)_{W1}$	19	19	0		
$(3,9)_{W2}$	14	14	0		
$(4,6)_{W1}$	10	7	3		
$(4,6)_{W2}$	5	4	1		
$(5,3)_{W1}$	12	5	7		
$(5,3)_{W2}$	3	3	0		
$(6,0)_{W}$	4	0	1	3	
$(3,1)_{W}$	6	0	6		

References

Blount, Sally (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes,* 63 (2), 131–144.

Bolton, Gary E., Jordi Brandts and Axel Ockenfels (2005). Fair procedures: evidence from games involving lotteries. *Economic Journal*, 115 (506), 1054–1076.

Bolton, Gary E. and Axel Ockenfels (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90 (1), 166–193.

Buchanan, James (1964). What should economists do? *Southern Economic Journal* 30: 213–222.

Cappelen, Alexander W., Astri D. Hole, Erik Ø. Sørensen, and Bertil Tungodden (2007). The pluralism of fairness ideas: An experimental approach. *American Economic Review*, 97 (3), 818–827.

Cappelen, Alexander W., James Konow, Erik Ø. Sorensen and Bertil Tungodden (2013). Just luck: an experimental study of risk-taking and fairness. *American Economic Review* 103(4): 1398–1413

Charness, Gary and Matthew Rabin (2002). Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics* 117: 817–869.

Dworkin, Ronald (1981). What is equality? Part 2: Equality of resources. *Philosophy and Public Affairs,* 10: 283–345.

Falk, Armin and Urs Fischbacher (2006). A theory of reciprocity. *Games and Economic Behavior* 54: 293–315.

Fehr, Ernst and Klaus Schmidt (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics* 114: 817–868.

Hayek, Friedrich (1976). *Law, Legislation and Liberty. Vol 2: The Mirage of Social Justice.* Chicago: University of Chicago Press.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat and Vernon L. Smith (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7 (3), 346–380.

Hoffman, Elizabeth and Matthew L. Spitzer (1985). Entitlements, rights, and fairness: An experimental examination of subjects' concepts of distributive justice. *The Journal of Legal Studies*, 14 (2), 259–297.

Isoni, Andrea, Anders Poulsen, Robert Sugden and Kei Tsutsui (2014). Efficiency, equality and labelling: an experimental investigation of focal points in explicit bargaining. *American Economic Review* 104: 3256–3287.

Mollerstrom, Johanna, Bjørn-Atle Reme and Erik Ø. Sørensen (2015). Luck, choice and responsibility: an experimental study of fairness views. *Journal of Public Economics* 131: 33–40.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review*, 90 (4), 1072–1092.

Levine, David (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics* 1: 593–622.

McCabe, Kevin, Mary Rigdon and Vernon Smith (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organization* 52: 267–275.

Milgram, Stanley (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67 (4), 371–378.

Rabin, Matthew (1993). Incorporating fairness into game theory and economics. *American Economic Review* 83: 1281–1302

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir (1991). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *American Economic Review* 81: 1068–1095.

Rawls, John (1971). *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.

Roemer, John E. (1998). *Equality of Opportunity*. Cambridge, Mass.: Harvard University Press.

Ruffle, Bradley J. (1988). More is better, but fair is fair: tipping in dictator and ultimatum games. *Games and Economic Behavior*, 23 (2), 247–265.

Sugden, Robert (2004). Living with unfairness: the limits of equality of opportunity in a market economy. *Social Choice and Welfare* 22: 211–236.

Zimbardo, Philip (2007). *The Lucifer Effect: Understanding How Good People Turn Evil*. New York: Random House.

Online Appendix: Instructions for Experiment

Welcome to today's experiment and thanks for coming. This is an experiment in decision-making. At the end of the experiment you will be paid the earnings you obtained from this experiment plus a participation fee of £3.

It is important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your cooperation.

I will now describe the nature of the tasks within the experiment.

Tasks

This experiment contains two parts. At the beginning of this experiment, individuals with odd seat numbers will become participant As and individuals with even seat numbers will become participant Bs. Each participant A will be randomly matched with a coparticipant B. This matching will stay the same throughout the experiment. You will never be told who your coparticipant is.

During the experiment, you and your coparticipant will compete for 12 lottery tickets numbered 1 to 12. At the end of the experiment, the experimenter will put 12 tickets with the numbers 1 to 12 on them into a bag. One of you will be asked to come forward and pick one ticket from the bag. The number on this ticket will be the number of the winning ticket. If you hold the winning ticket, you will get £24. If your coparticipant holds the winning ticket, he or she will get £24.

Part 1

[For Real Effort treatment]

In this part, you will be given a task and your coparticipant will be given the same task. You and your coparticipant will do the task independently. After you both have finished the task, your score will be compared with your coparticipant's score. **At the end of the task, the winner will get 9 lottery tickets and the loser will get 3 lottery tickets.**

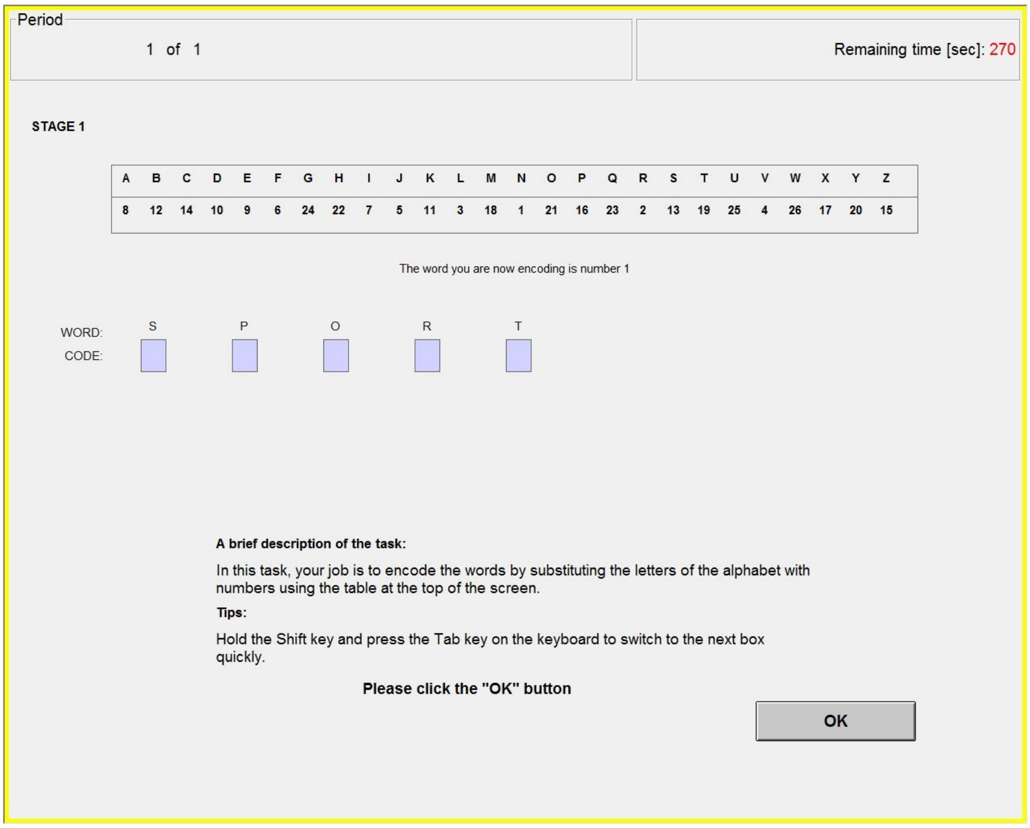
In the task, you will be presented with a number of words and your task will be to encode these words by substituting the letters of the alphabet with numbers using Table 1 below.

Table 1

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

Example 1: You are given the word FLAT. The letters in Table 1 show that F=6, L=3, A=8, and T=19.

In the task, Table 1 will also be shown on each screen. The picture below shows you how the computer screen will look.



All the codes need to be entered into the boxes under the letters of the word that you are asked to encode. You can shift among boxes by clicking the boxes. After you encode a word, you need to click the ‘OK’ button to verify your codes. The computer will tell you whether the word has been encoded correctly or not. If the word has been encoded wrongly, you need to check your codes and correct them. Then, you need to click the ‘OK’ button again to verify the codes.

Once you encode a word correctly, the computer will prompt you with another word which you will be asked to encode. Once you encode that word, you will be given another word and so on. This process will continue for 6 minutes (360 seconds).

You and your coparticipant will be given the same words to encode in the same sequence.

After you both have finished the task, the number of words you have encoded will be your score for the task. Your score will be compared with your coparticipant's score. If you and your coparticipant get the same score, then the computer will compare the total amount of time that you used encoding these words (i.e. the time between the start of the task and when the OK button was clicked after you finished the last word) with the total amount of time that your coparticipant used.

At the end of the task, the person with the higher score will be the winner. If you and your coparticipant get the same score, then the person who encodes the words in the shorter time will be the winner. The winner will get 9 lottery tickets and the loser will get 3 lottery tickets.

Please raise your hand if you have any questions.

Before you start to take decisions, we ask you to answer some questions in the next several screens. The purpose of these questions is to check whether you have understood these instructions. Any mistake you may make in doing these questions will not affect your final money earnings.

When you have finished Part 1, please remain seated. When everyone has finished Part 1, I will distribute the instructions for Part 2.

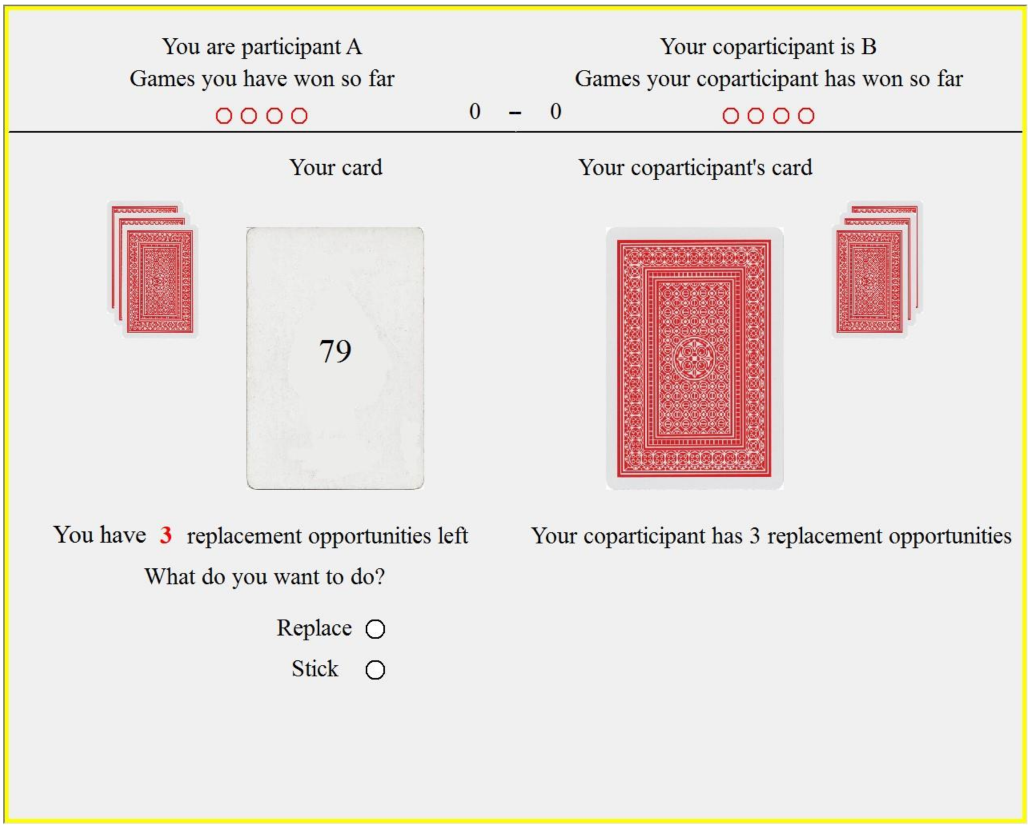
[For Fair Rule treatment]

In this part, you will play a series of card games with your coparticipant. **The winner of the series of card games will get 9 lottery tickets and the loser of the series of card games will get 3 lottery tickets.**

At the start of each card game, you will be dealt a card and your coparticipant will also be dealt a card. Each card has a number of points, which can be any of the whole numbers in the range from 1 to 100. Each of these numbers is equally likely at each 'deal'. You and your coparticipant will have some opportunities to replace cards. To win the game, you need to hold a card with a higher number of points than the card hold by your coparticipant.

You can decide whether to **stick** with the card you have been dealt or **replace** it with a new card. If you decide to replace the card with a new card, then the computer will randomly draw a new one for you. Each number in the range from 1 to 100 is still equally likely at this 'deal'. However, if you decide to replace this card, you cannot go back to it again.

In each game, if you are participant A, you are allowed to replace cards up to 3 times. If you are participant B, you are allowed to replace cards up to 3 times. During the game, you can decide to stick with any card that is dealt to you. Once you have used up all these replacement opportunities, you will not be able to make any further replacement and have to stick with the last card that you have been dealt. On the computer display, there will be a message reminding you how many replacement opportunities you have left. The picture below shows you how the computer screen might look in the first game, before you had made any decision.



While you are making your decisions about whether to stick or to use replacement opportunities, you will not know what decisions your coparticipant is making. Nor will you know the numbers on the cards that are dealt to him or her. Similarly, your coparticipant will not know what decisions you are making, or the numbers on the cards that are dealt to you.

After you and your coparticipant have made the decision of sticking with a card you have been dealt or have used up all your opportunities for replacing cards, your coparticipant’s card will be turned over. You can observe the points on your card and the points on your coparticipant’s card. You will also be shown how many replacement opportunities your coparticipant has used. Whoever has the card with the higher number of points on it wins the game. If you both have the same number of points, the game is a draw. The first participant to win 4 games will be the overall winner of the series of games. Draws will not be counted. The overall winner will get 9 lottery tickets and the overall loser will get 3 lottery tickets.

Please raise your hand if you have any questions.

Before you start to take decisions, we ask you to answer some questions in the next several screens. The purpose of these questions is to check whether you have understood these instructions. Any mistake you may make in doing these questions will not affect your final money earnings.

When you have finished Part 1, please remain seated. When everyone has finished Part 1, I will distribute the instructions for Part 2.

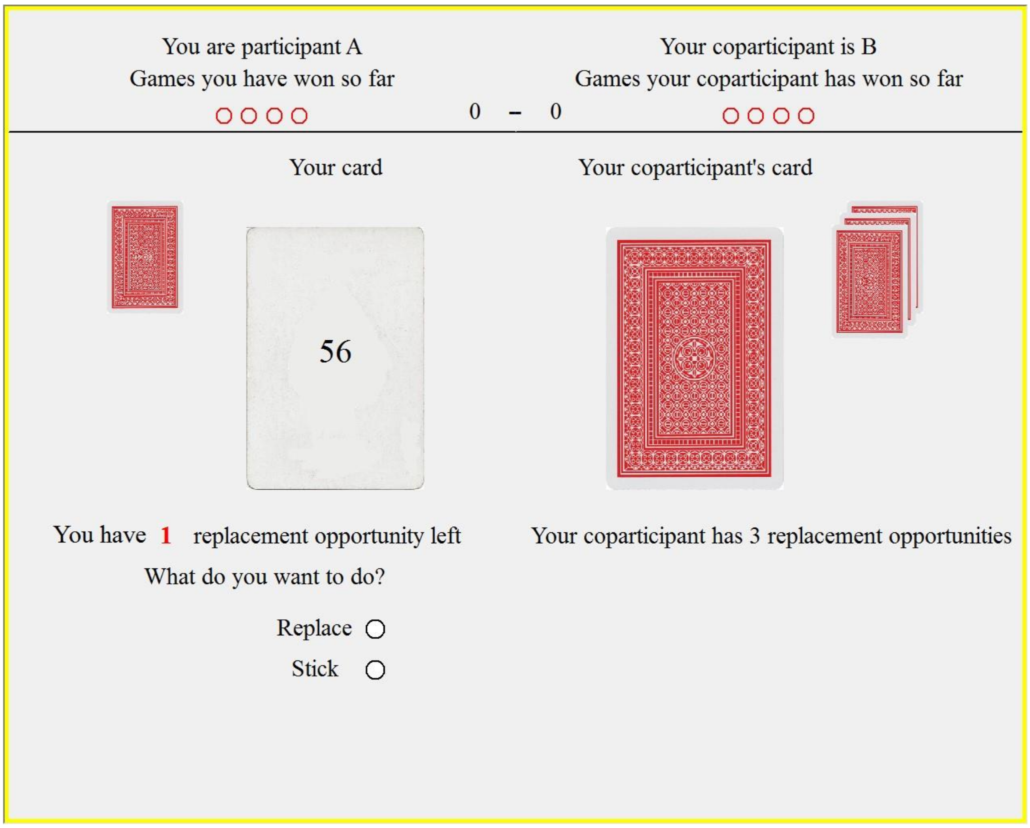
[For Unfair Rule treatment]

In this part, you will play a series of card games with your coparticipant. **The winner of the series of card games will get 9 lottery tickets and the loser of the series of card games will get 3 lottery tickets.**

At the start of each card game, you will be dealt a card and your coparticipant will also be dealt a card. Each card has a number of points, which can be any of the whole numbers in the range from 1 to 100. Each of these numbers is equally likely at each ‘deal’. You and your coparticipant will have some opportunities to replace cards. To win the game, you need to hold a card with a higher number of points than the card held by your coparticipant.

You can decide whether to **stick** with the card you have been dealt or **replace** it with a new card. If you decide to replace the card with a new card, then the computer will randomly draw a new one for you. Each number in the range from 1 to 100 is still equally likely at this ‘deal’. However, if you decide to replace this card, you cannot go back to it again.

In each game, if you are participant A, you are allowed to replace cards up to 1 time. If you are participant B, you are allowed to replace cards up to 3 times. During the game, you can decide to stick with any card that has been dealt to you. Once you have used up all these replacement opportunities, you will not be able to make any further replacement and will have to stick with the last card that you have been dealt. On the computer display, there will be a message reminding you how many replacement opportunities you have left. The picture below shows you what you might see on the computer screen in the first game, if you were participant A and before you had made any decision.



Participant B would see a similar screen, showing the card that he or she had been dealt and saying that he/she had 3 replacement opportunities left. While you are making your decisions about whether to stick or to use replacement opportunities, you will not know what decisions your coparticipant is making. Nor will you know the numbers on the cards that are dealt to him or her. Similarly, your coparticipant will not know what decisions you are making, or the numbers on the cards that are dealt to you.

After you and your coparticipant have made the decision of sticking with a card you have been dealt or have used up all your opportunities for replacing cards, your coparticipant's card will be turned over. You can observe the points on your card and the points on your coparticipant's card. You will also be shown how many replacement opportunities your coparticipant has used. Whoever has the card with the higher number of points on it wins the game. If you both have the same number of points, the game is a draw. The first participant to win 4 games will be the overall winner of the series of games. Draws will not be counted. The overall winner will get 9 lottery tickets and the overall loser will get 3 lottery tickets.

Please raise your hand if you have any questions.

Before you start to take decisions, we ask you to answer some questions in the next several screens. The purpose of these questions is to check whether you have understood these instructions. Any mistake you may make in doing these questions will not affect your final money earnings.

When you have finished Part 1, please remain seated. When everyone has finished Part 1, I will distribute the instructions for Part 2.

Part 2

At the end of Part 1, 12 lottery tickets were allocated between you and your coparticipant based on the result of the tasks you carried out. The winner in Part 1 got 9 lottery tickets and the loser in Part 1 got 3 lottery tickets. The tickets are numbered 1 to 12. The computer has picked 9 of these numbered tickets at random and assigned them to the winner. The remaining 3 tickets have been assigned to the loser. One of these lottery tickets will be the winning ticket, which gives a prize of £24. At the end of Part 2, the number of the winning ticket will be picked at random. Therefore, each lottery ticket gives a 1/12 chance of winning the prize. At the end of Part 2, if you hold the winning ticket, you will get the prize. If your coparticipant holds the winning ticket, he or she will get the prize.

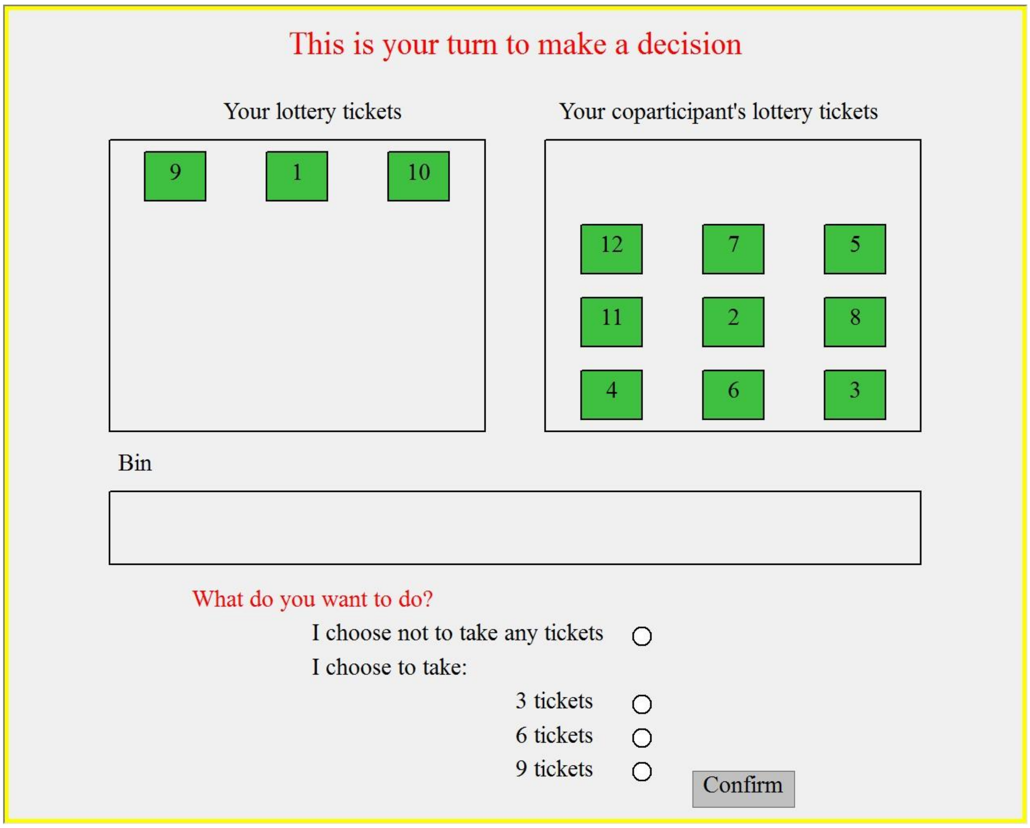
In this part of the experiment, you and your coparticipant will take turns to choose whether to stay with the current distribution of lottery tickets or to change it. The loser in Part 1 will make a choice first.

On the screen there will be three baskets. One contains the lottery tickets that you currently hold, one contains the lottery tickets that your coparticipant currently holds, and one is a bin.

When it is your turn to choose, you will be asked to decide whether you want to take some lottery tickets from your coparticipant, and if so, how many lottery tickets to take. Amounts taken have to be in blocks of three (so if you choose to take tickets, the number you take can be 3, 6 or 9, up to as many as your coparticipant has at the time). You always have the option of not taking any tickets.

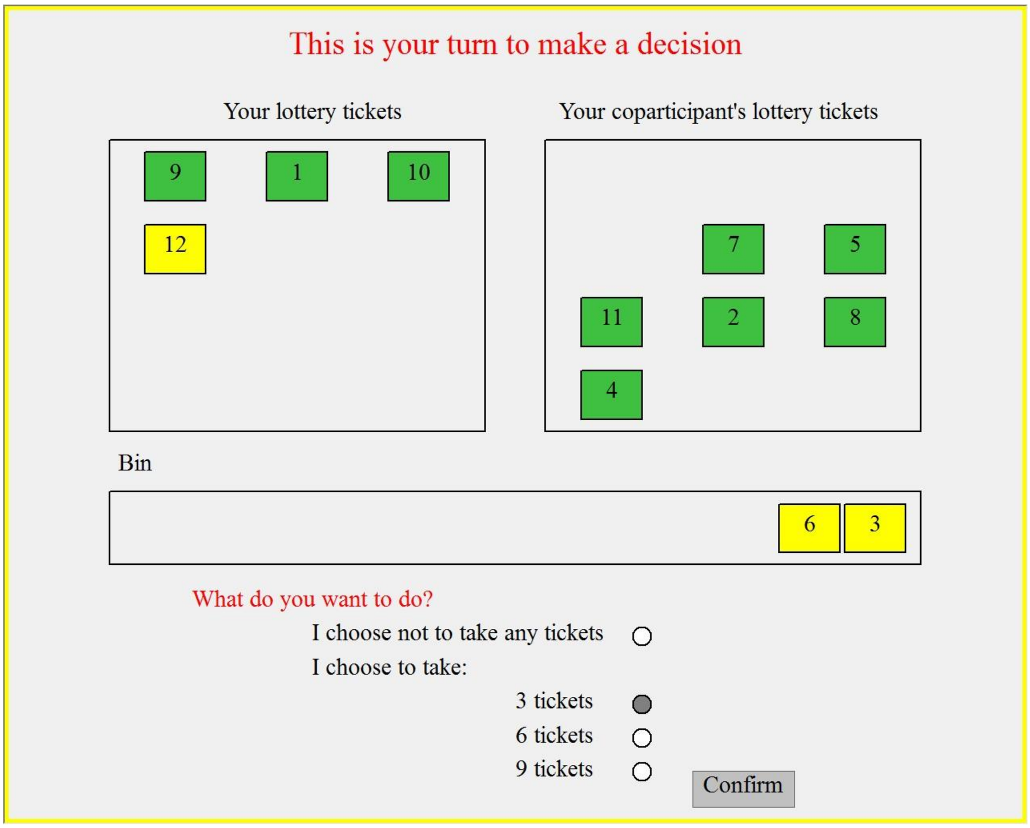
For every block of 3 lottery tickets that you take away from your coparticipant, one ticket from the block will be moved into your basket and the other two will be moved into the bin. If at the end of Part 2 the winning ticket is in the bin, neither you nor your coparticipant gets the prize.

The picture below shows how the computer screen would look at the start of Part 2 if you had been the loser in Part 1. Your basket is on the left, containing the three tickets that you earned in Part 1. Your coparticipant's basket is on the right, containing the nine tickets that he or she earned in Part 1. The bin is at the bottom. At the top of the screen you are told that it is your turn to make a decision.



Before making any decision, you are allowed to try different possible numbers of blocks of lottery tickets to take away from your coparticipant. The computer will show you how many of these lottery tickets will be moved from your coparticipant's basket into your basket and how many of these lottery tickets will be moved from your coparticipant's basket into the bin. These lottery ticket will be shown in a different colour.

For example, if you clicked the option 'I choose to take 3 tickets', the picture below shows you what the computer would display. From the picture, you can see that three lottery tickets have been moved from your coparticipant's basket. One of these (number 12) has been moved into your basket. The other two (numbers 3 and 6) have been moved into the bin. All these tickets (numbers 3, 6 and 12) are in yellow.



After your decision is made, you need to click the ‘Confirm’ button. The baskets and bin will be updated to show you the outcome of your decision and the current location of all the lottery tickets. All the lottery tickets will come back to being coloured green. Then it will be your coparticipant’s turn to make decisions on whether to take lottery tickets from you, and if so, how many lottery tickets to take. After your coparticipant has chosen, the baskets and bin will be updated again to show you the current location of all the lottery tickets as a result of your coparticipant’s choice. Then it will be your turn to choose again, and so on.

If there are four turns in a row (two for you and two for your coparticipant) in which neither of you takes lottery tickets, then Part 2 will end. Because tickets can be taken only in blocks of three, Part 2 will also end if your basket and your coparticipant’s basket both contain less than three tickets.

The experimenter will then put 12 numbered tickets into a bag. One of you will be asked to come forward and pick one ticket from the bag. The number on this ticket will be the number of the winning ticket. You will see whether this winning ticket is in your basket, or in your coparticipant’s basket, or in the bin. If the winning ticket is not in the bin, whoever holds it will get the prize of £24. If the winning ticket is in the bin, then neither you nor your coparticipant gets the prize. In all cases, both of you will also get a £3 participation fee.

Please raise your hand if you have any questions.

Before you start to take decisions, we ask you to answer some questions in the next several screens. The purpose of these questions is to check whether you have understood these instructions. Any mistake you may make in doing these questions will not affect your final money earnings.

When you have finished Part 2, please remain seated until everyone has finished Part 2.