Individual Sense of Fairness: An Experimental Study

By Edi Karni, Tim Salmon and Barry Sopher

The Johns Hopkins University, Florida State University and Rutgers University

August 2001

Abstract

This paper presents an experimental test of the theory of individual sense of fairness of Karni and Safra (2000) using a modified 3-player dictator game. The dictator is asked to allocate chances to win a single indivisible $15 dollar prize among himself and two others. His choice is restricted to a chord in the probability simplex. If the dictator chooses an interior point along the chord, this involves giving up own probability to win in exchange for a fairer allocation procedure. The results indicate that a strong preference for fairness exists in some subjects but not others. The chords used in the experiment were also constructed to allow the investigation of other properties of the subjects preferences for fairness.

Key Words :
fairness, dictator game

1. Introduction

By and large neoclassical economics is founded on a narrow notion of self-interest seeking behavior. Specifically, individual choice behavior is presumed to be motivated solely by the desire to attain higher levels of *material well-being*. This presumption stands in stark contrast to long held views in philosophy and psychology maintaining that human behavior is motivated in part by emotions and, in particular, by moral sentiments[1]. Recently, however, there is growing interest among economists in the potential implications of broadening the psychological base of the model of individual behavior by incorporating emotions into the theory of choice. (See, for example, a survey by Elster [1998] and discussions by Loewenstein [2000], Romer [2000].) This interest is partly due to experimental evidence of cooperative behavior in situations in which maximization of material self-interest alone would imply non-cooperative behavior.[2] In this context, emotions play the role of enforcement mechanisms.

In this paper we explore, via experiments, some issues pertaining to the presence of an intrinsic sense of fairness as a motive force in individual choice behavior. Specifically, confronting subjects with choices among allocation procedures that involve random selection of a winner of a predetermined prize, we test for evidence of a willingness to sacrifice one's own chance of winning to attain what is perceived to be a fairer allocation procedure. This work is inspired by two recent papers of Karni and Safra (2000, 2000a), the first of which contains an axiomatic model of choice behavior that is motivated, in part, by concern for fairness and the second of which introduces measures of the intensity of the sense of fairness and derives their behavioral characterizations. Our experimental

1 See Hume (1740), Smith (1756), Rawls (1963, 1971) for philosophical discussions.
2 For instance, Camerer (1997), Berg, Dickhaut, and McCabe (1995).

design is based on the analytical framework of Karni and Safra (2000). Our results, therefore, constitute a direct test of their contention concerning the manifestation of inherent sense of fairness in individual behavior.

To set the stage we summarize briefly those elements of the theory of Karni and Safra that are relevant for our work. In doing so we focus on aspects of the axiomatic theory that underlie the experimental design and on aspects of the measurement of the intensity of the sense of fairness that are relevant for the interpretation of our findings.

Let $N = \{1, ..., n\}, n > 2$, be a set of individuals who must decide on a procedure by which to allocate, among themselves, one unit of an indivisible good. Clearly, the ex-post allocation is necessarily unfair. One individual is awarded the good and the others are not. The issue, therefore, is what allocation procedure may be implemented to attain a higher level of fairness ex-ante. Karni and Safra (2000) restrict attention to procedures that allocate the good by lot. Formally, denote by e^i, the unit vector in R^n, the *ex-post* allocation in which individual i is assigned the good. Let $X = \{e^i \mid 1 \leq i \leq n\}$ be the set of *ex-post* allocations and let P be the $n-1$ dimensional simplex representing the set of all probability distributions on X. In this context P has the interpretation of the set of *random allocation procedures*.

Each individual is represented by two binary relations on P, the relation $\succ$, representing his actual choice behavior, and the relation $\succ_F$, representing his concept of fairness. The relation $\succ$ has the usual interpretation, namely, for any pair of allocation procedures p and q in P, $p \succ q$ means that, if he were to choose between p and q, the individual would choose p or would be indifferent between the two. The fairness relation, $\succ_F$,has the interpretation of "fairer than." In other words, $p \succ_F q$ means that the allocation pro-

cedure p is regarded by the individual as being at least as fair as the allocation procedure q. In general, the notion of fairness may be subjective or objective. In either case, it is intrinsic and, jointly with concern for self-interest, governs the individual's choice behavior among allocation procedures.

Taking the preference and the fairness relations as primitives, Karni and Safra (2000) state conditions that permit the derivation of the self-interest motive implicit in the individual choice behavior. Loosely speaking, an allocation procedure p is preferred over another allocation procedure q from a self-interest point of view if the two allocation procedures are equally fair and p is preferred over q. Moreover, Karni and Safra introduce axioms that are equivalent to the existence of an *affine* function $\kappa : P \to \mathsf{R}$ representing $\prec_S$, the derived binary relation representing the *self-interest component* of the preference relation $\prec$, a strictly quasi-concave function $\sigma : P \to \mathsf{R}$ that represents the fairness relation $\prec_F$, and a utility function V representing the preference relation $\prec$ as a function of its self-interest and fairness components, i.e., for all allocation procedures, $p, q \in P$,

$$p \prec q \Leftrightarrow V\left(\left(\kappa \cdot p, \sigma\left(p\right)\right)\right) \geq V\left(\left(\kappa \cdot q, \sigma\left(q\right)\right)\right).$$

In addition, Karni and Safra (2000) examine the case in which the function V is additively separable in the self-interest and fairness components. Formally,

$$V\left(\left(\kappa \cdot p, \sigma\left(p\right)\right)\right) = h\left(\kappa \cdot p\right) + \sigma\left(p\right),$$

where h is a monotonic increasing function.

The experimental design used to test this theory is a three person version of a dictator game in which the dictator must choose how to allocate the chances of winning the prize. The dictator is not given complete freedom to pick any allocation he desires, rather he must select the allocation from a predetermined set of such allocations represented by a

chord in the probability simplex.

Studies of three person ultimatum games include Güth and Van Damme (1998) and Bolton and Ockenfels (1998,1999). Typically these games involved one person proposing a split among all three players with one of the other two being designated to accept or reject the proposed split. In our experiments, there is no acceptance/rejection decision. Their proposal was that people care only about the average payoff to the other two and find that players seem relatively unconcerned with the distribution among the other players. Kagel and Wolfe(1999) examine three person ultimatum games allowing for a "consolation prize" to the third party if a proposal is rejected to examine some reciprocity issues.

Charness and Rabin (2000) investigate a wide variety of games including a few three person dictator games in an attempt to distinguish between models of fairness as resulting from "difference aversion" as proposed by Fehr and Schmidt (1999), or from what Charness and Rabin (2000) calls "quasi-maximin" preferences as proposed by Andreoni and Miller (2000). Individuals who possess "difference averse" preferences dislike having any individual's payoff too different from any others. Persons with "quasi-maximin" preferences are less interested in any absolute differences in payoffs, but are more interested in helping out those players with low payoffs than those with higher payoffs.

The three person dictator games in Charness and Rabin (2000), however only allow dictators to make binary choices which restricts the potential richness of choices that more options along a chord will allow. Their primary interest appears to be in investigating reciprocity issues and not necessarily fairness as such, though. Their results do show that subjects are willing to give up some potential for gain in an attempt to equalize the

payoff to the other subjects, much as our results will show. Engelmann and Strobel (2001) present the results from similar experiments in three person dictator games in which the subjects choices are also limited to binary choices. They, however, designed the available choices to test between specific models of fairness. They find some evidence in favor of difference averse preferences but also find that there are substantial deviations from such a model that may be best explained through efficiency or "maximin" considerations.

Our method is similar to the experiments of Andreoni and Miller (2000). Andreoni and Miller have subjects allocate coins between themselves and another subject along particular exchange rates. This is equivalent to presenting the subjects with a choice from multiple possible budget sets. Their interest was in determining the degree to which subjects would violate the basic notions of revealed preference theory. Our design has a similar interpretation: the chord in the probability simplex is the equivalent of a budget set and the person's choice represents the point along that budget constraint that intersects with the "highest" indifference curve. Our purpose, however, is different. We are interested in examining the nature of those indifference curves and, in particular, whether they indicate a willingness to trade one's own chance of winning for a higher degree of fairness. We do, however, include some questions in our experiments to examine the consistency of the subjects choices.

The rest of the paper is organized as follows: In Section 2 we describe the design of the experiments. In section 3 we discuss in more detail the theoretical foundations of the measures of preferences for fairness we will use. In Section 4 we present and analyze the findings. The main conclusions and issues raised by this work are summarized in Section 5.

2. The Experiments: Design and Implementation

The experiment described below is intended to test the degree to which subjects are willing to trade-off their own chance of winning to attain a fairer lottery that will determine the winner of an indivisible good. The design of these experiments is a three person dictator game and will use a modified version of the interface used in Sopher and Narramore (2000). The mechanism for making a choice in the two settings is the same, but the contexts of the choices are significantly different.

The experiments involved bringing groups of subjects in multiples of three into a computer lab. The subjects were given a verbal introduction to the experiment including an overview of the rules and they were then lead through an interactive help program to make sure that they understood the interface and rules of the experiment. When the subjects had completed the instructions sequence, each subject is randomly assigned a type of either A, B or C. The subjects were then divided anonymously and randomly into three person groups with one subject of each type in each group. The player A in each group was asked to choose the allocation of the probabilities to the subjects in the group that will be used in the actual lottery for a $15 prize. Put differently, subjects of type A are asked to design a lottery $p = (p_A, p_B, p_C)$, where $p_i \geq 0$, $i = A, B, C$ and $\Sigma_{i=1}^{3} p_i = 1$, to be used to select the winner of the $15. The question to the players A was phrased as follows: "Please choose the allocation of chances to be used in deciding who among A, B, and C wins the prize."

The B and C players in each group are asked to make similar choices but their choices did not affect their payoffs. The B players were asked to respond to the question: "Please select the allocation that you would choose if you were the decision maker, Player A."

The players C were asked to respond to the question: "Please select the allocation that you believe is fair." The main purpose of doing this is to give the other players a choice task such that no player can identify who is a player A based upon observing some players making a choice and others not. As a secondary consideration, though, it is of interest to examine the preferences expressed by the other players in a hypothetical context, and to elicit their views on what the fair allocation procedure is.

The A players were informed that they were the only ones making a choice affecting the payoffs of everyone in their group and the B and C players were fully informed that their choices would not effect their payoffs. The A players were further informed that the B and C players will be responding to other questions, but they were not told what those other questions were. For precise information on what the subjects were told, there is a complete record of the help screens that the subjects are lead through to explain the experiment contained in the Appendix.

The choice set corresponding to this design is a 2-dimensional simplex depicted in Figure 1. The top vertex of the triangle represents the allocation procedure that gives the entire probability of winning to the subject of type A. The lower left vertex is the allocation procedure that gives the entire probability to subject of type B and the one at the lower right to subject of type C. The intensity of the subjects sense of fairness can be represented in this context by the curvature in their indifference curves.[3] At one extreme, if the subject of type A exhibits no sense of fairness (i.e., is only concerned about his own chance of winning), then his indifference curves will be straight lines, such as line 1, along which p_A is constant. If, however, he is concerned about the fairness of

[3] See Karni and Safra (2000a) for a detailed analysis.

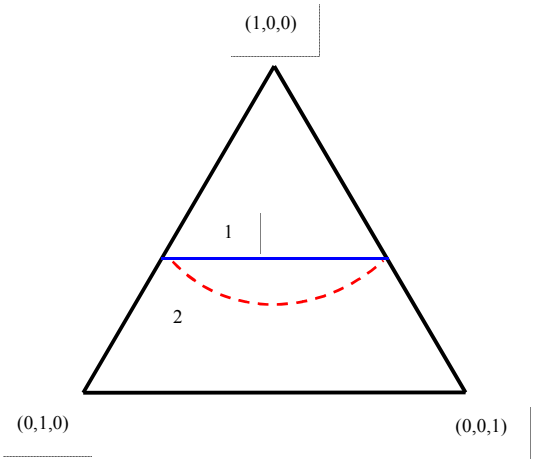

Figure 1: Characterization of indifference curves. Line 1 characterizes a person with no preference for fairness. Line 2 characterizes a person who exhibits a preference for fairness.

the allocation procedure and, in particular, if he regards the two other subjects in the group as equally deserving of the prize, then the indifference curves may be convex as shown by line 2 indicating strict preference for fairness. This indifference curve depicts a willingness to sacrifice one's own probability of winning to attain a fairer allocation procedure. In Section 3 we discuss possible shapes of indifference curves that can represent preferences for fairness in more detail.

In the experiment, a subject is presented with a chord in this simplex along which his own probability of winning varies with the probabilities of the other subjects, and is asked to choose a point along it. If the subject is not concerned about fairness then he should select the endpoint that gives him the maximum probability to win the prize. If, however, the subject does possess a sense of fairness, and if fairness calls for the assignment of equal probabilities of winning to other subjects that have equal claims for the prize (i.e., equal treatment of equals) then the optimal lottery may be represented by a point in the interior of the chord.

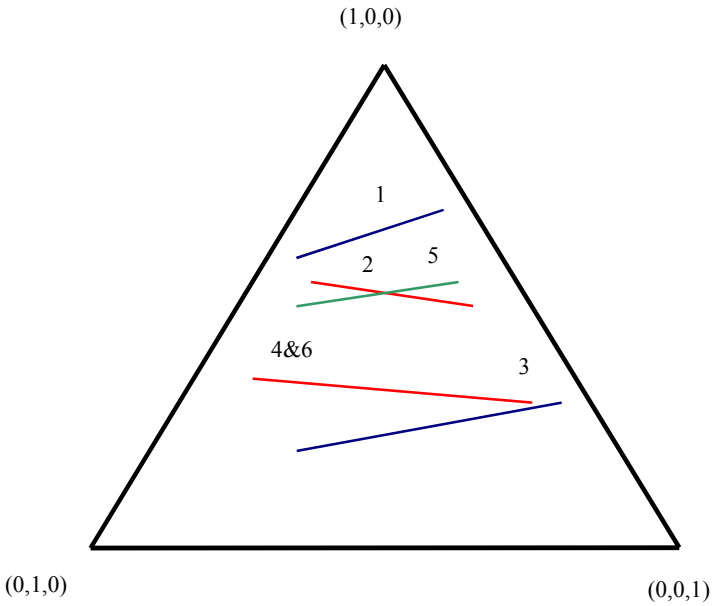

Figure 2: Graphical representation of the chords used in the experiment.

The subjects were asked to make a total of six choices along 5 different chords. These are depicted in Figure 2 and the lotteries defining the endpoints of the chords are shown in Table 1. The chords chosen possessed some specifically designed similarities to allow the investigation of particular issues and this will be discussed in more detail later. After each choice, the groups of players were reshuffled randomly, but the subjects retained their type throughout the experiment. Thus, a subject who was assigned type A at the outset remains type A for all six trials.

The existence of multiple sequential choices raises the possibility that subjects could engage in behavior based upon compounding the lotteries across choices. This could have lead an A player to think that he was being fair to his counterparts by staying at the starting endpoint on all of the choices. The reasoning would be that over the course of the experiment, this might equalize the chance of winning for players B and C. Consequently, even if an A player did not move from the endpoint, it would not have been possible to

	Endpoint 1			Endpoint 2		
	p_{A}	p_{B}	p_{C}	p_{A}	p_{B}	p_{C}
$Q1$	70	5	25	60	35	5
$Q2$	55	35	10	50	10	40
$Q3$	30	5	65	20	55	25
$Q4\&Q6$	35	55	10	30	10	60
$Q5$	55	10	35	50	40	10

Table 1: Lotteries defining the endpoints of the chords used in the experiments.

conclude that this choices were not motivated, in part, by concern for fairness. To address this issue, only the first lottery was used to generate actual payoffs. The choices made by the A types on this question were used in the actual lotteries that determine the winner of the \$15 prize. To ensure that the subjects believed the lotteries were run fairly, an extra subject was recruited in each session to run the lotteries with a pair of 10-sided dice and then observe that the proper amount of money was inserted into envelopes to pay the subjects at the end of the experiment.

Since only one chord was used to run an actual lottery, this raises a question concerning the reliability of the answers given for the other five questions. To aid in determining the degree to which this is important, half of the subjects were asked question 1 first and the other half question 2. By checking the degree of consistency between the choices of the two groups on the paid and unpaid question, we can test the degree to which subjects display a stronger preference for fairness when the decision is hypothetical versus when it is real.

An example of the interface used in this experiment can be found in the Appendix. The subjects were presented with an initial allocation indicating the chances out of 100 for each subject in the group to win the lottery. These chances appeared as colored slices of a pie. Subjects can use a slider bar to move along the chord between this point and the

other endpoint. With each movement of the slider bar, both the chances of the subjects to win and the pie chart were updated accordingly. Their final choice can be represented by a number $\lambda \in [0, 1]$ such that λ is the weight used to create the convex combination of the endpoints resulting in the chosen allocation. For all questions, a choice of $\lambda = 1$ indicates that the player A chose the point that maximized his probability of winning while a choice of $\lambda = 0$ indicates that player A chose an allocation procedure that minimized his chance of winning.

One complication to this analysis was introduced due to the desire to list the probabilities of each subject winning the lottery as integers between 0 and 100. This caused the actual chords the subjects were choosing along to be jagged instead of smooth. The formulae used to generate these probabilities were :

$$\begin{aligned} p_A &= Round(\lambda \overline{p}_A + (1-\lambda)\bar{q}_A) \\ p_B &= Round(\lambda \overline{p}_B + (1-\lambda)\bar{q}_B) \\ p_C &= 100 - p_A - p_B \end{aligned} \tag{1}$$

where p_i is the probability allocated to player $i \in \{A, B, C\}$ by the choice of λ while $\bar{p}_i$, is the probability that player i would win at the upper endpoint and $\bar{q}_i$ is the probability at the lower endpoint. The slider bar used had 30 discrete "click" points along it that the subjects could choose. λ was then calculated by taking the "click" point along the slider bar chosen and dividing it by the number of discrete clicks that were made available.

The subjects used in these experiments were drawn from two separate subject pools. One group of subjects consisted of (mainly) undergraduate and (some) graduate students at The California Institute of Technology (CIT), and the other consisted of students from Pasadena City College (PCC). In total there were 135 subjects used in these experiments

with 69 coming from the CIT population and 66 coming from the PCC population. Each session run contained subjects from only one group or the other.

Earnings from these sessions consisted of 1 out of every 3 subjects winning a $15 prize in addition to their show-up fee and the other 2 out of 3 subjects receiving only their show-up fee. For CIT subjects, the show-up fee was $5 and for PCC subjects the show-up fee was $10[4]. The sessions for these experiments lasted as few as 20 minutes and one up to 40 minutes. Most finished in between 20 and 30 minutes.

3. Preference Structures

In this section we describe the possible effects of alternative notions of fairness on the behavior of the subjects in the experiments. These examples are intended to help interpret the empirical findings and to highlight the role of the concept of fairness in the design of the optimal allocation procedures. We adopt, for the this purpose, the additively separable version of Karni and Safra (2000). Thus, the utility function of a subject i can be decomposed into a fairness component represented by a function $\sigma_i(p)$, and a self-interest component represented by a function, $h_i(\kappa \cdot p)$ where $p = (p_A, p_B, p_C)$. There are infinitely many potential candidates to use for the functional forms of these relations, corresponding to different notions of fairness. Here we only consider the two which we describe as "equal treatment of equals," (ETE) and "equal treatment of others" (ETO).

[4] The reason for the differential is simply to encourage PCC students to travel the extra distance to Caltech where the experiments were run. In addition, some subjects redeemed recruitment coupons worth $10 that are given to PCC students when they sign-up to be on the recruitment list to be used in their first experiment.

3.1. Fairness as equal treatment of equals

The first notion of fairness calls for impartial treatment of individuals whose claims for the prize are deemed of equal merit. This notion is similar in spirit to the idea of "difference aversion" as proposed in Fehr and Schmidt (1999). We capture this idea by the assumption that the fairness of a lottery is determined by its proximity to the center of the simplex or the lottery $p = (1/3, 1/3, 1/3)$. Formally, assume that

(A.1) For all subjects, the fairness relation is symmetric with indifference curves that are concentric circles and a utility representation that is proportional to the distance from the center of the simplex. Formally,

$$\hat{\sigma}_i(p) = -\hat{\theta}_i \sqrt{\sum_{j \in \{A,B,C\}} \left(p_j - \frac{1}{3}\right)^2}$$

where i denoted the name of the subject, p_j, $j = A, B, C$ is the probability type j winning, and $\hat{\theta}_i > 0$ is an individual parameter for person i.

The symmetry assumption seems justified on the grounds of "the principle of insufficient reason." In each experiment the participating subjects came from similar backgrounds and they are unaware of the identity of the other members of their group. Thus, they possess no information that would allow them to apply some other ethical consideration to differentiate among the claims of the different subjects in the group to the prize. Put differently, it seems reasonable to assume that the claims of all subjects in each group to the prize are of equal merit and should be treated impartially. Thus, from behind this 'veil of ignorance' each lottery (p_A, p_B, p_C) is equally fair, or equally unfair, as any of its permutations. The assumption regarding the shape of the indifference curves and the specific functional form are chosen for convenience.

The utility representation of the self-interest component is described in assumption (A.2).

(A.2) For all subjects the self-interest component depends solely on the own probability of winning. Formally,

$$\hat{h}_i\left(\hat{\kappa}\cdot p\right)=\hat{\kappa}_i p_i$$

with $\hat{\kappa}_i>0$.

The presumption underlying (A.2) is that there are no within-group externalities in the enjoyment of the prize.

A chord in our experiment is defined by two lotteries, $\bar{p}=(\bar{p}_A,\bar{p}_B,\bar{p}_C)$ and $\bar{q}=(\bar{q}_A,\bar{q}_B,\bar{q}_C)$, that represent its endpoints in the simplex. The chord itself corresponds to the subset $T\left(\bar{p},\bar{q}\right)=\{\lambda\bar{p}+(1-\lambda)\bar{q}\mid\lambda\in[0,1]\}$ consisting of all the lotteries that can be obtained as a convex combination of $\bar{p}$ and $\bar{q}$. Assume throughout that $\bar{p}_A>\bar{q}_A$ and let $\widehat{\lambda}_A^f\left(\bar{p},\bar{q}\right)=\arg\max_{\lambda\in T(p,q)}-\hat{\theta}_i\left[\sum_{j\in\{A,B,C\}}\left(\lambda\bar{p}_j+(1-\lambda)\,\bar{q}_j-\frac{1}{3}\right)^2\right]^{1/2}$. Then, $p\left(\widehat{\lambda}_A^f\left(\bar{p},\bar{q}\right)\right)=\widehat{\lambda}_A^f\left(\bar{p},\bar{q}\right)\bar{p}+\left(1-\widehat{\lambda}_A^f\left(\bar{p},\bar{q}\right)\right)\bar{q}$ represents the lottery along the given chord that is deemed the fairest. Under these assumptions

$$V_i(\lambda\bar{p}+(1-\lambda)\,\bar{q})=\hat{\kappa}_i\cdot(\lambda\bar{p}_i+(1-\lambda)\,\bar{q}_i)-\hat{\theta}_i\left[\sum_{j\in\{A,B,C\}}\left(\lambda\bar{p}_j+(1-\lambda)\,\bar{q}_j-\frac{1}{3}\right)^2\right]^{1/2}$$

is concave in λ. Hence, we have:

Proposition 1 *Let $\hat{\lambda}_i^*\left(\bar{p},\bar{q}\right)$ be the optimal choice for person i of type A of an allocation procedure in $T\left(\bar{p},\bar{q}\right)$. If assumptions (A.1) and (A.2) hold then $\hat{\lambda}_i^*\left(\bar{p},\bar{q}\right)$ is unique and:*

1. $\hat{\theta}_i/\hat{\kappa}_i\leq\frac{(p_A-q_A)\left[\sum_{j\in\{A,B,C\}}\left(p_j-\frac{1}{3}\right)^2\right]^{1/2}}{\sum_{j\in\{A,B,C\}}\left(p_j-\frac{1}{3}\right)(p_j-q_j)}=\hat{\xi}$ implies $\hat{\lambda}_i^*\left(\bar{p},\bar{q}\right)=1$,
2. $\hat{\theta}_i/\hat{\kappa}_i>\frac{(p_A-q_A)\left[\sum_{j\in\{A,B,C\}}\left(p_j-\frac{1}{3}\right)^2\right]^{1/2}}{\sum_{j\in\{A,B,C\}}\left(p_j-\frac{1}{3}\right)(p_j-q_j)}=\hat{\xi}$ implies $\hat{\lambda}_i^*\left(\bar{p},\bar{q}\right)\in[\hat{\lambda}_A^f\left(\bar{p},\bar{q}\right),1)$.

The expression $\hat{\theta}_i/\hat{\kappa}_i=\hat{\phi}_i$ is a measure of the intensity of the sense of fairness.[5] It may be assumed to be distributed over the half-open interval $[0,\infty)$. Given a chord $(\bar{p},\bar{q})$ this induces a distribution on $\widehat{\lambda}_i^*\left(\bar{p},\bar{q}\right)$. Because the actual range of $\widehat{\lambda}_i^*\left(\bar{p},\bar{q}\right)$ is $[\widehat{\lambda}_A^f\left(\bar{p},\bar{q}\right),1]$ and the effect of censoring the induced distribution of $\widehat{\lambda}_i^*\left(\bar{p},\bar{q}\right)$ tends to

[5] See Karni and Safra (2000a) for a general treatment.

	$\hat{\xi}$	$\widehat{\lambda}_{\mathsf{A}}^{\mathsf{f}}(\bar{p},\bar{q})$
$Q1$	0.45	0.25
$Q2$	0.19	0.48
$Q3$	0.16	0.37
$Q4,6$	0.07	0.55
$Q5$	0.19	0.48

Table 2: Critical values of $\hat{\phi}$ for each question along with the values of λ that induce the most fair allocations according to the ETE hypothesis.

have a concentration at 1. Moreover, one implication of the theory is that the probability of observing $\widehat{\lambda}_i^*(\bar{p},\bar{q}) < \widehat{\lambda}_A^f(\bar{p},\bar{q})$ is zero.

Clearly, given $\hat{\theta}_i$ and $\hat{\kappa}_i$, $\widehat{\lambda}_i^*(\bar{p},\bar{q})$ and the maximal fairness position $\widehat{\lambda}_A^f(\bar{p},\bar{q})$ depend on the particular chord. Applying the proposition to the questions in the experiments we get the values seen in Table 2.

Observe that the critical values $\hat{\xi}$ and $\widehat{\lambda}_A^f(\bar{p},\bar{q})$ are the same in $Q2$ and $Q5$ reflecting the fact that questions 2 and 5 are permutations of each other in which the probabilities of players of type B and C are switched. Notice also that the range of $\widehat{\lambda}_i^*(\bar{p},\bar{q})$is larger for $Q1$ than it is for $Q3$. (The same observation and latter observations as well apply to $Q2$ and $Q4$.) Recall that chords 1 and 3 have the same slope. This is a manifestation of the "wealth effect" mentioned in section 2. To state this idea more formally, let $\bar{p}$ and $\bar{p}'$ in P be such that $\sigma(\bar{p}) = \sigma(\bar{p}')$, $(\bar{p}' - \bar{p})(0,1,-1) = 0$, and $\bar{p}'_A - \bar{p}_A > 0$ (see Figure 3). For $\bar{q}', \bar{q} \in P$ we say that $T(\bar{p}',\bar{q}')$ *chord-dominates* $T(\bar{p},\bar{q})$ if $T(\bar{p}',\bar{q}')$ and $T(\bar{p},\bar{q})$ have the same slope and the same length. The range $[\hat{\lambda}_A^f(\bar{p},\bar{q}),1]$ is monotonic increasing with respect to the chord-dominance relation (i.e., $\hat{\lambda}_A^f(\bar{p}',\bar{q}') \leq \hat{\lambda}_A^f(\bar{p},\bar{q})$.) Moreover, it can be shown that under ETE, an individual willingness to sacrifice his own probability of winning to bring about a fairer allocation procedure is monotonic increasing with respect to chord-dominance. Formally,

Proposition 2 *Under the ETE assumptions (A.1) and (A.2), if* $T(\bar{p}', \bar{q}')$ *chord-dominates* $T(\bar{p}, \bar{q})$ *then* $\hat{\lambda}_i^*(\bar{p}, \bar{q}) \geq \hat{\lambda}_i^*(\bar{p}', \bar{q}')$

The proof is immediate and is not given here. However, the logic of Proposition 2 is that a subject's own probability of winning is a component of the allocation procedure and, as such, has an impact on the overall fairness of the procedure. Thus, when this probability is high, giving up some contributes, in itself, to the fairness of the procedure while if it is low giving up the own probability of winning detracts from the fairness of the procedure. In the former case, the sacrifice involved in giving up the own probability of winning is mitigated by the contribution it makes to increasing fairness (this is in addition to the gain in fairness attained by reallocation of the probabilities between the other two subjects). In the latter case either the sacrifice involved in giving up the own probability of winning outweighs the effect of the increase in fairness, or is exacerbated by the diminution of fairness, it causes. Thus, in general, the higher is the own probability of winning the larger is the mitigating and the smaller is the exacerbating force, which together give rise to the principle of increasing benevolence. (This logic is illustrated in Figure 3.)

This is an example of the more general principle that follows:

The Principle of Increasing Benevolence: *The willingness to sacrifice one's own probability of winning to bring about a fairer allocation procedure increases with respect to chord-dominance.*

3.2. Fairness as equal treatment of others

Another possible way subjects might view issues of fairness is in terms of the treatment of only others in the group. According to this view, the fairest allocation procedure along

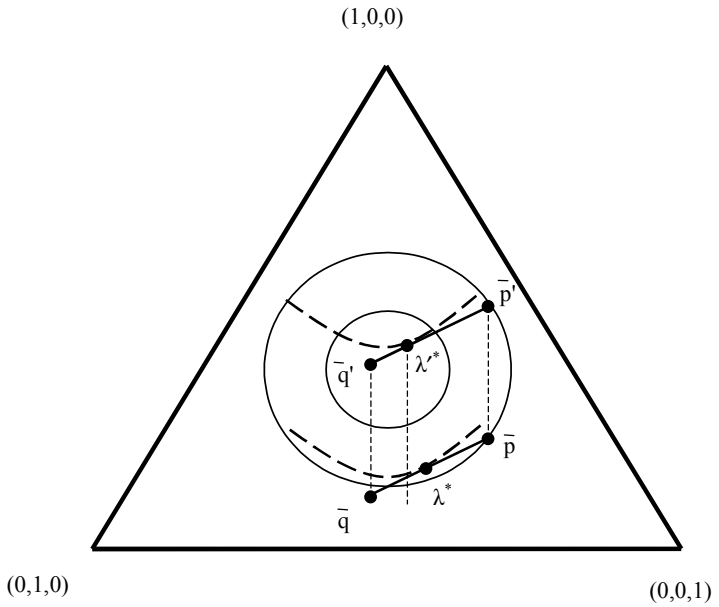

Figure 3: Graphical representation of the principle of increasing benevolence as predicted by the ETE notion of fairness. Notice that the optimal choice on the upper chord is at a point farther down the chord.

a given chord is the one that minimizes the difference in the probabilities of the other two subjects in the group. Subjects may be willing to give up some of their own probability of winning to reduce the gap between the probabilities of the other members in the group. In some ways, this specification of preferences for fairness is a specialization of the idea of quasi-maximin preferences discussed in Charness and Rabin (2000) to this context. Formally, assume A.2 and:

(B.1) For any allocation procedure p, and any three subjects, i, k, l, subject i's fairness relation is represented by

$$\sigma_i(p) = -\theta_i(p_k - p_l)^2$$

where p_k and p_l are the winning probabilities of subjects k and l.

If individual i is of type A then the combination of these two assumptions results in a utility function as follows: Given two endpoints of a chord p and q, the utility of the

lottery implied by a choice of λ is

$$V_i(p,q,\lambda) = \kappa_i(\lambda\bar{p}_A + (1-\lambda)\,\bar{q}_A) - \theta_i((\lambda\bar{p}_B + (1-\lambda)\,\bar{q}_B) - (\lambda\bar{p}_C + (1-\lambda)\,\bar{q}_C))^2$$

Consequently, the choice of λ that results in the lottery deemed most fair, $\lambda_A^f\,(\bar{p},\bar{q})$, is simply the λ that induces $p_B = p_C$.

As in the case of equal treatment of equals, this utility function implies indifference curves of the sort shown in Figure 1. In both instances, if $\theta_i = 0$, then the indifference curves are straight lines indicating that the individual is only concerned about their own probability of winning. If $\theta_i > 0$, the indifference curves will be curved as line 2 in Figure 1 indicating that lotteries that lead to more equal splits between the other two subjects are preferred to those that lead to less equal splits.

Assuming that the optimal λ has been chosen, λ^*, we can solve for θ_i/κ_i which gives us the value for our measure of intensity of preference for fairness, ϕ_i: implied by that choice of λ.

$$\phi_i = \frac{\theta_i}{\kappa_i} = \frac{(\bar{p}_A - \bar{q}_A)}{2\,(\lambda^*\bar{p}_B + (1-\lambda^*)\,\bar{q}_B - \lambda^*\bar{p}_C - (1-\lambda^*)\,\bar{q}_C)\,(\bar{p}_B - \bar{q}_B - \bar{p}_C + \bar{q}_C)} \tag{3.2}$$

To find the critical value, ξ, of ϕ_i such that any value of $\phi_i < \xi$ results in an observed choice of $\lambda = 1$, we can just set $\lambda^* = 1$, in equation 3.2. The critical values for each question as well as the values of λ that result in the most fair outcome for this notion of fairness are contained in Table 3.

A variation of the hypothesis of equal treatment of others assume that, for any allocation procedure p, subject i's fairness relation is represented by

$$\sigma_i\,(p) = -\theta_i\,|p_k - p_l|$$

	ξ	$\lambda_{\mathsf{A}}^{\mathsf{f}}(\bar{p},\bar{q})$
$Q1$	0.50	0.60
$Q2$	0.18	0.54
$Q3$	0.09	0.33
$Q4,6$	0.06	0.52
$Q5$	0.18	0.54

Table 3: Critical values of ϕ for each question along with values for λ that lead to the allocations deemed most fair according to the ETO hypothesis.

and his self interest is represented by A.2.

From the point of view of subject of type A the point of maximal fairness is the same as in the previous version of the model but the indifference curves are V shaped (i.e., consist of two linear segments) and symmetric about the maximal fairness point. Hence, for any given chord the utility function is linear in λ and its slope depends on ϕ. Consequently, type A's optimal position is either $\lambda_A^*(\bar{p},\bar{q}) = 1$ if the slope of the indifference curve is smaller than that of the chord or $\lambda_A^*(\bar{p},\bar{q}) = \lambda_A^f(\bar{p},\bar{q})$ (i.e., at the point along the chord where $p_B = p_C$) if the slope of the indifference curve is steeper than that of the chord. Thus, for each chord this model implies a bimodal distribution with concentrations at 1 and $\lambda_A^f(\bar{p},\bar{q})$.

Remark: Note that the principle of increasing benevolence does not apply when the fairness relation is based on the principle of "equal treatment of others." This is because the own probability of winning does not Figure in the fairness relation. This observation is used to test the hypothesis of equal treatment of others.

4. Results and Analysis

4.1. Methods

To answer the questions posed in the previous sections, we will employ several modes

of analysis. Some will consist of straightforward distribution tests and require little explanation. We will also derive several of our results from a more carefully constructed ordered probit analysis and this requires a bit of explanation.

The experimental results consist of choices that are probability mixtures $(\lambda, (1-\lambda))$ of two lotteries, where, for each question, λ denotes the weight on the lottery that is most favorable to player A. Underlying these choices, we suppose, is the subject's weighting of the importance of the fairness of the overall allocation procedure relative to his or her own probability of winning, as outlined in Section 2. Let $\lambda_i \in [0,1]$ be the observed choice of subject i, from the discrete choice set $\{\alpha_0 = 0, \alpha_1, \alpha_2, ..., \alpha_{J-1}, \alpha_J = 1\}$ and assume that $0 < \alpha_1 < \alpha_2 ... < \alpha_{J-1} < 1$. Let $\widetilde{\phi}_i = \mathsf{x}_i\beta + \varepsilon_i$ be an unobservable latent random variable measuring the subject i's intensity of the sense of fairness and ε_i is a random variable representing unobservable factors determining the subject's choice. This is a noisy approximation of the ϕ or $\widehat{\phi}$ described in Section 2. Then there exist threshold parameters, $(\mu_1, ..., \mu_J)$, such that if $\mu_j \leq \phi_i \leq \mu_{j+1}$ then $\lambda_i = \alpha_j$, for all $j = 0, 1, ..., J-1$. The only covariates x_i that we have are 0-1 dummy variables for different questions and conditions (e.g., whether the question was for real money or not). Thus, in the estimation we will mainly be making inferences about the true value of ϕ. The ordered model involves both estimated coefficients on the covariates (the β_i), as well as the unknown threshold parameters, μ_j. In general, both the estimated threshold parameters, μ_j, and the estimated β_i 's give us information about the true value ϕ_i. The probability that $\lambda_i = 0$ is $\Pr(\lambda_i = \alpha_0 = 0) = \Pr(\phi_i \leq \mu_1) = \Pr(\mathsf{x}_i\boldsymbol{\beta} + \varepsilon_i \leq \mu_1) = F_{\varepsilon_i}(\mu_1 - \mathsf{x}_i\boldsymbol{\beta})$ where F_{ε_i} denotes the cumulative distribution function of ε_i. Similarly, for the other outcomes, $\Pr(\lambda_i = \alpha_j) = \Pr(\mu_j \leq \mathsf{x}_i\boldsymbol{\beta} + \varepsilon_i < \mu_{j+1}) = F_{\varepsilon_i}(\mu_{j+1} - \mathsf{x}_i\boldsymbol{\beta}) -$

$F_{\varepsilon_i}(\mu_j - \mathsf{x}_i\boldsymbol{\beta})$ where $j = 0, 1, ..J$, $\mu_0 = -\infty$, $\mu_j \leq \mu_{j+1}$,and $\mu_{J+1} = \infty$.

Note, however, that for a given notion of fairness, if the fairest allocation, $\lambda^f(\bar{p}, \bar{q})$ (or $\widehat{\lambda}^f(\bar{p}, \bar{q})$, we will use $\lambda^f(\bar{p}, \bar{q})$ to refer to both in this section) is in the interior of the interval $[0, 1]$ then the strict prediction of the theory is that $\Pr(\lambda < \lambda^f(\bar{p}, \bar{q})) = 0$. This is a testable implication of the theory. In fact, for the specific hypotheses ETE and ETO we do observe a few choices of $\lambda^* < \lambda^f$, implying that for some subjects the strict versions of these hypotheses are falsified. If we allow, however, for the possibility that some individuals may express their preferences imprecisely, then these hypotheses may still be valid, provided the implied error rate is not too implausible. It is, of course, possible that some subjects' idea of fairness are different from either ETE or ETO in which case the observed choices of λ do not violate the general theory even though it violates the specific hypotheses concerning the notion of fairness.

The likelihood function is constructed in the usual way from these probabilities. If the ε_i's are $N(0, 1)$ then we have the ordered multinomial probit model, and the maximum likelihood estimator can be derived in the usual manner, using first-order conditions for the maximum of the likelihood function. Note that if we have no covariates, then, under the assumption that the $\varepsilon_i's$ are standard normal, $\Pr(\lambda_i = \alpha_j|\boldsymbol{\mu})) = \Pr(\mu_j \leq \varepsilon_i < \mu_{j+1}) = \Phi(\mu_{j+1}) - \Phi(\mu_j)$. Thus, the estimates of the threshold parameters are simply the numbers that replicate, via the standard normal distribution, the observed frequencies of choice in each of the possible "cells". This does not provide any direct information about the value of ϕ, the intensity of the sense of fairness, but it does admit of the following interpretation. Since the $\varepsilon_i's$ are meant to represent what we, the observers, do not know about the value of ϕ for individuals in our sample, the higher choices (near to 1) are

interpreted as being driven by relatively smaller values of ϕ (a weaker sense of fairness). That is, larger ε_i's are associated with a smaller ϕ, and vice-versa. The addition of 0-1 indicator variables have a similar interpretation, though their precise meaning depends upon the case. In general,

$$\Pr(\lambda_i = \alpha_j) = \Pr(\mu_j \leq \mathsf{x}_i\boldsymbol{\beta} + \varepsilon_i < \mu_{j+1}) = \Phi(\mu_{j+1} - \mathsf{x}_i\boldsymbol{\beta}) - \Phi(\mu_j - \mathsf{x}_i\boldsymbol{\beta}).$$

In the estimated model we include indicators for each distinct question in the experiment, for whether the question was played out for real money, from which of two subject pools the subject came from, and whether or not the subject was male. For any of the indicator variables, the estimated coefficients indicate a shift in the whole distribution to either the right or the left, depending on whether the sign of the coefficient is positive or negative. For the question dummy variables, this shift should be interpreted as due to factors that differ among the structure of questions. For the indicator variable for whether the question was played out for real money, the interpretation of a positive (negative) coefficient estimate would be that there is a decrease (increase) in the displayed intensity of the sense of fairness, perhaps due to a shift in relative emphasis on one's own payoff and the overall fairness of the resulting distribution of payoffs engendered by consideration of the fact that one's choice will have a real impact on one's own, and others, well-being. In the case of the indicator variables for subject pool and sex, the interpretation of a positive (negative) coefficient is that the average intensity of the sense of fairness is lower (higher) for the subject pool or gender that the indicator stands in for (i.e., "Caltech" or "Male").

Table 4 contains results of the ordered probit regression models (one for each subject type) which help us to summarize efficiently the within- and between-question differences. Since choices are ordered along each chord in the simplex, we can treat each

Indep Variable	A's	B's	C's
MALE	.48 (1.84)*	.04 (0.15)	.33 (1.52)
CIT	.39 (1.53)	.58 (2.40)**	-.14 (-0.62)
PAY	.41 (3.51)***	.01 (0.03)	-.06 (-0.39)
Q2	-.38 (-2.87)***	.16 (0.71)	-.45 (-2.73)***
Q3	-.13 (-0.69)	-.27 (-1.21)	-.31 (-1.44)
Q4	-.46 (-3.31)***	.05 (0.19)	-.51 (-2.24)**
Q5	.07 (.37)	-.19 (-0.78)	-.14 (-.70)
Q6	-.38 (-2.77)***	-.08 (-0.30)	-.48 (-2.02)**
N	270	270	270
LL	-626.96	-528.21	-671.91
PRsq	.03	.02	.01

Table 4: Ordered Probit Results for Choice Mixture Chosen. z-statistics in parentheses. *, **, or *** to indicate significance at the 10, 5 or 1% level.

mixture choice with positive mass in the distribution of choices as a discrete choice. Table 4 summarizes results for ordered probit regressions for each of the three subject types. The dependent variable is CHOICE, the mixture chosen, λ_i, while the regressors are MALE (1 if male, 0 if female), CIT (1 if Caltech student, 0 if PCC student), PAY (1 if a paid question, 0 otherwise), and Q2-Q6 indicators for questions 2 through 6. The coefficients and corresponding z-statistics are reported for each variable. The Table also reports N (number of independent observations), LL (log-likelihood for the estimated model), and PRsq (pseudo-R-squared) for each regression. The z-statistics are computed with adjusted standard errors based on an estimator of variance independently developed by Huber (1967) and White (1980) to take account of the fact that there are multiple observations (6) on each individual.

The signs and significance of the coefficients can be interpreted as outlined above. In general, a positive (negative) coefficient indicates either a weaker (stronger) intensity of the sense of fairness (e.g., for the MALE variable), or a factor that leads subjects to change their intensity of sense of fairness (e.g., the PAY variable). The coefficients of the

dummy variables for the various questions may be interpreted as the effect of the position of the corresponding chord in the simplex. Given the underlying concept of fairness, this position has an effect of changing the λ that corresponds to the maximal fairness and, consequently, the optimal allocation procedure as well as the level of concentration due to the effect of censoring. Overall, there are more significant effects of the different indicators for the type A subjects then type B subjects. (The type C subjects appear to act similarly to type A subjects.) The only significant effect for type B subjects is for the subject pool variable, indicating that, on average, the Caltech subjects acted more selfishly, but there were no differences between questions.

While the signs of the coefficients indicate the general nature of the shift in behavior, a more precise measure of the effects of the variables can be obtained by computing the marginal effects of the variables on the estimated probabilities of the various choices. Since there are between 31 and 36 discrete choice values in the three regressions, it is neither practical nor interesting to compute marginal effects for all possible choices. We compute marginal effects for the modal choices (λ =.5333 (about 18% of all choices) and $\lambda = 1$ (about 33% of all choices), which account for about one half to two thirds of choices over the different subject types. While the signs of the coefficients in the regression results reported in Table 4 do not necessarily imply a particular direction of change for the probability of a given choice option being chosen, since the two choices we consider are at opposite end of the distribution and are the modal choices, the effect is apparent: a positive sign indicates an increase in the probability that $\lambda = 1$ and a decrease in the probability that $\lambda = .5333$. A negative sign indicates the reverse. The effect on the probability that intermediate values of λ are chosen is ambiguous, in general, though

Type/Choice	Type A		Type B		Type C	
	$\lambda = .533$	$\lambda = 1$	$\lambda = .533$	$\lambda = 1$	$\lambda = .533$	$\lambda = 1$
MALE	-.08*	.17*	-.01	.02	-.03	.11
CIT	-.06	.14	-.07**	.23**	.01	-.04
PAY	-.07***	.16***	.00	.00	.01	-.02
Q2	.06***	-.13***	-.02	.06	.03**	-.13***
Q3	.02	-.05	.03	-.11	.03	-.09*
Q4	.07***	-.15***	-.01	.02	.03**	-.14***
Q5	-.01	.02	.02	-.08	.01	-.04
Q6	.06**	-.13***	.01	-.02	.03**	-.13**
% of Choices	19%	33%	14%	54%	20%	26%

Table 5: Marginal Effects on Probability of Choice for Modal Choices. *, **, or *** to indicate significance at the 10, 5 or 1% level.

there will be, necessarily, a rightward shift in the distribution from a positive effect and a leftward shift in the distribution for a negative effect.

Table 5 contains the results of these computations. The Table shows the change in the probability of choice for a change from 0 to 1 of each independent variable, computed with the other independent variables at their mean sample values. We do not show the z-statistics (which are very similar to those of the estimated coefficients in the regressions), but only indicate whether the effect is significant at the 1%, 5% or 10% level.

An example of how to read these results is that male type A subjects are 17% more likely to choose $\lambda = 1$ (and 8% less likely to choose $\lambda = .5333$) than female subjects, type B subjects from Caltech are 23% more likely to choose $\lambda = 1$ (and 7% less likely to choose $\lambda = .5333$) than are subjects from Pasadena City College, and type A subjects are 16% more likely to choose $\lambda = 1$ (and 7% less likely to choose $\lambda = .5333$) on paid questions than on unpaid questions. Table 6 provides an interpretation of the results in Table 5 for differences across the different questions, treating all insignificant effects as zero.

Question	Paid/Unpaid	$\lambda = .5333$	$\lambda = 1$
1	unpaid	baseline	baseline
1	paid	-7%	+16%
2	unpaid	+6%	-13%
2	paid	-1%	+3%
3	unpaid	0	0
4	unpaid	+7%	-15%
5	unpaid	0	0
6	unpaid	+6%	-13%

Table 6: Interpretation of Marginal Effects on Probability of Modal Choices, A Players Only. (All effects relative to baseline)

4.2. Results

There are six major questions proposed or implied by the discussions in the earlier parts of the paper. Each one will be dealt with individually.

4.2.1. Comparison of Choices Between Player Types

Figures 4-6 contain histograms of the answers given by the subjects to each of the questions asked in the experiment. Each Figure contains, for each question, three histograms, one for each subject-type. In each histogram, λ_A^f is identified by a dotted line and $\widehat{\lambda}_A^f$ by a solid line corresponding to the "fairest" choices according to the ETO and ETE notions of fairness, respectively. Since subjects were assigned a type for the entire experiment, the three histograms in each Figure represent independent samples.

Before proceeding, one, unanticipated, feature of the data must be noted. Subjects of type B were asked what they would choose if they were the type A player, the deciders. In designing their role we intended them to suppose themselves in the position of type A and make their choices accordingly. Subjects of type C were to choose what they believed to be the "fair" allocations. It seems, however, that some type B and C subjects answered the questions as if they were in the position of having the power to determine

the allocation procedure, but with themselves still occupying the place of player B or C, respectively, and getting the chances of winning for a player B or C from any given allocation. On chord choices that slope down to the left (1, 3 and 5), some type B subjects (12% of them) chose the lowest point on the chord, and on chord choices that slope down to the right (2, 4 and 6), some type C players (11% of them) also chose the lowest point of the chord. Such choices maximize their chances of winning the prize while minimizing the chances of player of type A to win the prize. This behavior makes no sense from the point of view of an A player, since the choice entails simultaneous sacrifice of the subject's own chances of winning and of the fairness of the allocation procedure as a whole. We interpret these choices as reflecting a misunderstanding of the point-of-view that they were supposed to take. To make our case we note that while occasionally subjects of type A chose an allocation procedure that minimized their own chances of winning, this occurred in less than 1% of choices. Consequently, in comparing the three histograms in each Figure, it is more instructive to shift the mass on the choice of $\lambda = 0$ in the relevant histogram to $\lambda = 1$. Both the histograms and the statistical tests reported below are performed on choice data with this shift performed.

With the foregoing adjustments made, the distributions are rather easily summarized. The typical pattern for all types of players is bimodal, with a mode in the neighborhood of, but not exactly at, .5, and a second mode at 1. On all of the questions, the distribution of type A and type C choices appear to be remarkably similar. The distributions for type B players are also, typically, bimodal, but with a larger mode at 1 than is the case for type A or C players. Thus, a mode at 1 of the distribution of choices by subjects of type A may be explained by the censoring effect discussed in section 2. Perhaps the most important

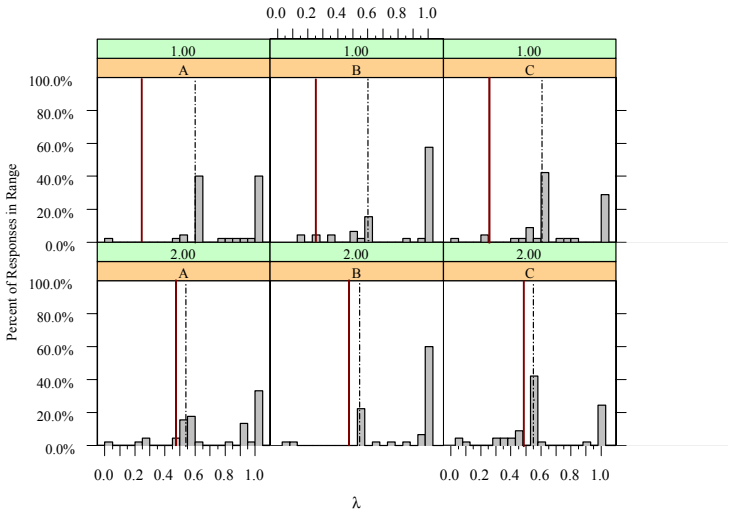

Figure 4: Histograms of choices by player type and by question for questions 1 and 2. Solid line $\widehat{\lambda}^f$ and dotted line represents λ^f.

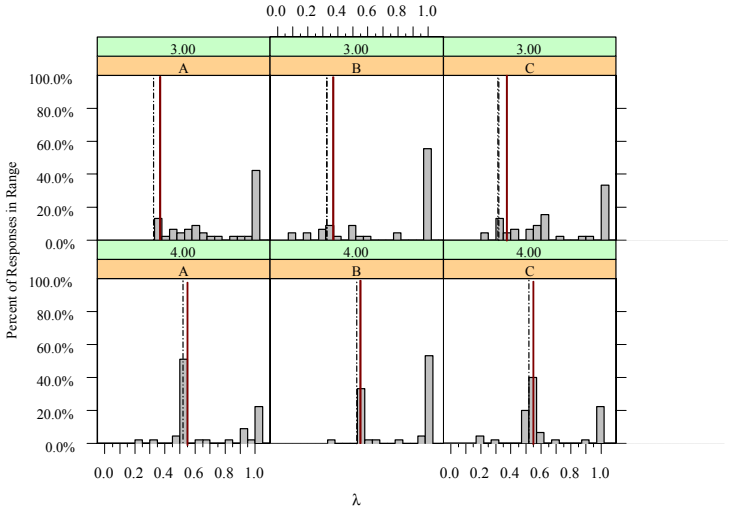

Figure 5: Histograms of choices by player type and by question for questions 3 and 4. Solid line $\widehat{\lambda}^f$ and dotted line represents λ^f.

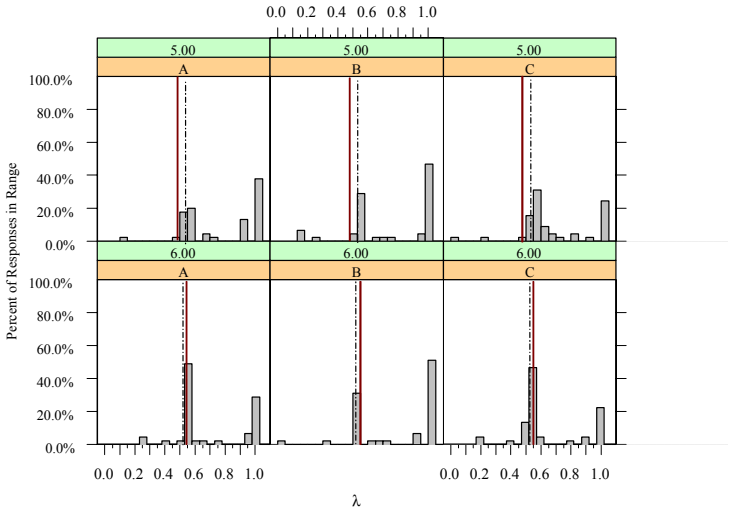

Figure 6: Histograms of choices by player type and by question for questions 5 and 6. Solid line $\widehat{\lambda}^f$ and dotted line represents λ^f.

finding regarding the answers of subjects of type A is the willingness of a substantial number of them to trade off their own probability of winning to attain a fairer overall allocation of these probabilities among the subjects in the group. The distribution of type B players choices suggest that subjects act differently when the are in the role of actually deciding the allocation procedure, as the type A's do, than when they are asked to choose but know that their choice will make no difference. The distribution of the answers of type C, being bimodal, as for the type A's, is something of a puzzle. It suggests that they entertain two distinct notion of fairness. While a significant number of type C subjects indicated a concept of fair allocation procedure similar to that articulated in section 2 above, a significant number of type C subjects believe that it is only fair that the person who has the power to decide takes full advantage of his position by assigning himself the maximum feasible probability of winning. They may, alternately, simply be placing

Question	All types	A and B	B and C	A and C
1	4.47 (.11)	-0.66 (.51)	1.93 (.05)	1.85 (.07)
2	15.82 (.00)	-2.42 (.02)	4.00 (.00)	1.99 (.05)
3	1.78 (.41)	-0.18 (.86)	1.13 (.26)	1.26 (.21)
4	13.83 (.00)	-2.79 (.01)	3.64 (.00)	1.31 (.19)
5	2.10 (.35)	0.09 (.93)	1.12 (.26)	1.44 (.15)
6	10.43 (.01)	-2.06 (.04)	3.28 (.00)	1.53 (.13)

Table 7: Tests for Differences in Choice Distributions Among Types (within Question). All types: Kruskal Wallis test, Chi-squared, 2 d.o.f., (p-value). Pairwise tests: Wilcoxon 2-sample test, standard normal z, (p-value).

themselves in the role of the A player and considering the trade-off between own and others' payoffs in their own preferences.

The apparent differences and similarities in the distributions for different subjects types noted above are substantiated for many of the questions by formal statistical tests. Table 7 contains the results of tests among the different subject types for each of the six questions using the data normalized as discussed above while Table 8 contains the results of the tests with the raw data. In both Tables, column 1 contains results of Kruskal-Wallis tests for the hypothesis that there is no difference between the distribution of choices for Types A, B and C. The Tables show the test statistic, which is Chi-squared with 2 degrees of freedom, and the associated p-value for the test in parentheses. Columns 2 to 4 contain results of Wilcoxon two-sample tests for the hypothesis that there is no difference in the distribution of choices for each possible pair of subject types. The Tables show the test statistic, which is a standard normal, with the associated p-value in parentheses.

Let us first look at the results with the normalized data, Table 7. Questions 2, 4 and 6 all show strong evidence of a difference overall. For each of these questions there turns out to be a strong pairwise difference between types A and B and between types B and C. For question 2 there is a significant difference between types A and C as well.

There is weaker evidence of an overall difference in type choices for Question 1 (11% significance level), apparently driven by differences between types B and C and between types A and C but not, as in the other cases, by a difference between types A and B. To summarize, there are more significant differences between types A and B (3 at the 5 % level), and between types B and C (4 at the 5% level) than between types A and C (1 at the 5 % level), consistent with the observation suggested by the histograms that Types A and C are more similar.

Due to the similarity of choice behavior between A's and C's, we might conclude that when someone is asked to make a decision, they tend to do so in a manner others consider fair. The fact that B's choose differently is quite interesting then. They were asked to select the choice they would make if they were in the position of player A. This choice is something of a double hypothetical choice since they are not actually choosing an allocation that will be used in the paid questions and on the unpaid questions, even player A is not making a "real" choice on most of the questions. Our intuition was that hypothetical questions would result in more generous or fair responses but in this case the B's choices are more selfish than those making the decision. One way of interpreting these results is that they are from a "bravado" like effect. An analogy would be people who are willing to suggest one course of action to others (i.e. a boss should fire an employee, a judge should convict a defendant) yet if actually given the responsibility to make the choice, they do otherwise (i.e. they decide not to fire employee or decide to acquit the defendant), perhaps because they consider more deeply the full ramifications of their decision, given that it is not hypothetical.

There is another novel interpretation to this in that it may shed some light on the

issue of the tendencies of expressive versus instrumental voters. Voters are described as expressive when they vote under the assumption that their vote will have no consequence for deciding the outcome and vote to express a view. Voters are described as instrumental when they vote believing that their vote will be decisive to the outcome. The conventional wisdom is that instrumental voters vote in a more self-interested manner than expressive voters. These results suggest otherwise. The B players in these experiments were making a choice, knowing it would have no consequences and might be described as similar to expressive voters. The A players were making a choice knowing it would affect the outcome. Our results show that the B players make more self-interested choices. The analogy between our choice environment and the voting context is not perfect, so this should only be taken as suggestive of what might happen in a voting context.

One thing that should be checked is the difference our normalizations lead to in the analysis and this can be seen in Table 8. Without the normalization, we always find that there is a significant difference across all types and between individual types A and B and then types B and C. The only tests that do not seem to result in significant differences are on question 1, 3 and 5 between types A and C. It should be noted that these were the questions that lead to C being advantaged at $\lambda = 1$. These results reinforce the differences between type B players and the other two types. They do lead to weaker similarities between A's and C's though on questions 2, 4 and 6 as these were the questions on which some subjects appeared to misinterpret the question they were being asked.

4.2.2. Paid versus Unpaid questions

Considering the fact that 5 of the 6 questions the subjects answered generated no payoffs, it is reasonable to ask what, if any, effect this had on the answers. A standard hypothesis

Question	All types	A and B	B and C	A and C
1	9.57 (.00)	2.92 (.00)	-1.91 (.06)	1.85 (.07)
2	51.89 (.00)	-2.42 (.02)	6.77 (.00)	5.43 (.00)
3	7.89 (.02)	2.69 (.01)	-1.92 (.06)	1.26 (.21)
4	39.26 (.00)	-2.79 (.01)	6.04 (.00)	4.30 (.00)
5	14.90 (.00)	3.68 (.00)	-2.72 (.01)	1.44 (.15)
6	38.69 (.00)	-2.06 (.04)	5.99 (.00)	4.76 (.00)

Table 8: Tests for Differences in Raw Choice Distributions Among Types (within Question). All types: Kruskal Wallis test, Chi-squared, 2 d.o.f., (p-value). Pairwise tests: Wilcoxon 2-sample test, standard normal z, (p-value).

from "induced value theory" (Smith (1976)) is that subjects will be more selfish when the choice has the possibility to generate a payment than when it will not. The experiment was set up to facilitate addressing this question by having half of the subjects see question 1 first and answer it knowing it will generate a payment and then having the other half answer question 2 first knowing it will generate a payment. The answers on these questions can be compared under paid and unpaid situations to determine if there is a systematic difference in behavior under the two treatments.

The first piece of evidence on this subject comes from the ordered probit results in Table 6 that shows that the type A subjects were 16% more likely to choose $\lambda = 1$, and 7% less likely to choose $\lambda = .533$, when the question generated a payment then when it did not. The B and C subjects displayed no such tendency.

Another way of testing this hypothesis is to test for differences in the distribution of choices on question 1 when it generates a payment versus when it does not and do the same for question 2. On question 2, the average λ chosen by the A players when it is paid is .82 versus .64 for when it is unpaid. For question 1 the average λ chosen by the A players is .77 for both paid and unpaid. Wilcoxon rank-sum tests to test for the equivalence of the distributions result in a Z statistic of -0.5411 and associated p-value

of. 0.5884 for question 1 and for question 2, the results are a Z statistic of -2.3615 and p-value of 0.0182. Consequently we find that on question 1, there is no statistically significant difference in the responses by the A players when the question is paid versus unpaid but there is at the 5% significance level for the responses to question 2. We note that since there were only 45 type A players, these tests are comparing a distribution of 22 choices to a distribution of 23 choices for each question. Overall, although the Wilcoxon tests do not uniformly support the estimated effect of pay from the ordered probit model, there is support for the idea that pay makes a difference, and that it tends to induce more selfish choices than when pay is not present.

4.2.3. Size of the Prize Effect

The chords used for questions 4 and 6 were identical with the only difference being the size of the hypothetical prize. On question 4, the prize was $15 while on question 6 the hypothetical prize was $45. The question is whether or not the size of the prize affects an individuals willingness to trade-off their own probability of winning to obtain a fairer allocation procedure.

The results from the ordered probit regression in Table 5 show that the marginal effects seen in the responses to questions 4 and 6 are close to identical, indicating that there is little difference in the responses obtained on the two questions. A Wilcoxon signed rank test results in a Z statistic of -0.2836 and p-value of 0.7767. Thus we find no statistically significant difference in the choices on these two questions that can be attributed to the size of the hypothetical prize. Note that this does not preclude the possibility that a larger real money prize could have an effect.

4.2.4. Symmetry

Symmetry requires that behavior on questions 2 and 5 should be similar as these chords are identical except the positions of the B and C players have been reversed. Thus if the type A players treat B and C symmetrically, their choices on the two chords should be similar if not identical.

The estimated marginal effects on the modal choices (.5333 and 1) based on the ordered probit results, shown in Table 6 , suggest a lack of symmetry between question 2 (unpaid) versus question 5 as there is more selfish behavior on question 5. Wilcoxon signed rank sum tests and Kolmogorov-Smirnov tests for the differences in the distribution of $\lambda's$ across questions 2 and 5 can be found in Table 9. These tests were performed on the raw choice data. These results show that for the A players, we can not detect a difference in the distribution of choices between these two questions. It is important to underscore the fact that these are the real deciders and therefore their behavior is of primary interest. We should also note that the reason the marginal effects suggest the symmetry property does not hold while the distribution tests show otherwise is that the distribution of choices for B and C players change dramatically between the two questions and this is the effect picked up in the marginal effect results.[6] Since these tests compare the entire distributions, rather than just the modes, we conclude that the A players do treat B and C symmetrically.

[6] As a note, if we test separately the choices made by A's who were paid based on question 2 and then test those who were not paid, the results from signed rank rum tests are a test statistic of 0.26 and p-value of 0.7942 for the subjects who were paid for question 2 and a test statistic of -1.68 with a p-value of 0.09 for those who were unpaid. In both cases we can conclude that there is no statistically significant difference in the choices at a 5% significance level, but it is something of a curiousity that the choices appear possibly more different when both questions are hypothetical than when only one is.

	A's	B's	C's
Wilcoxon Signed Rank Sum Test	-1.15 (0.25)	3.92 (0.00)	-3.96 (0.00)
Kolmogorov-Smirnov	0.089 (0.995)	0.444 (0.00)	0.622(0.00)

Table 9: Tests of the differences in the distribution of choices on questions 2 and 5 by each player type. The test statistic is listed with its associatated p-value in parentheses.

4.2.5. ETE versus ETO

The final issue to look at is to try to determine which of the two preference notions described in Section 3 best describe the data. There are two pieces of evidence to consider.

The first piece of evidence to examine is in the distribution of the $\lambda's$ chosen by the subjects in relation to the most fair $\lambda's$ associated with the two preference notions. An examination of the histograms in Figures 4-6 shows that in virtually all of the cells, there is a large mass of choices centering around λ_A^f, the most fair λ under the ETO notion, and that the choices by the subjects appear to trace its movements more closely than they track the movements of $\widehat{\lambda}_A^f$, the most fair λ under the ETE notion. Table 10 presents the results from one-sided Wilcoxon signed rank-sum tests that the distribution of $\lambda's$ are closer to λ_A^f than $\widehat{\lambda}_A^f$. The table shows that the choices of types A and C are highly statistically significantly closer to λ_A^f than $\widehat{\lambda}_A^f$ regardless of whether you consider all choices of λ or only those choices of $\lambda < 1$. For the B players considering only choices of $\lambda < 1$ the result is not statistically significant. The meaning of this result can be derived from what was noted at the end of Section 3.2 that if we use an alternative specification of the ETO hypothesis with the absolute value specification, then the prediction is that a bimodal result should be observed with all choices at either 1 or λ_A^f. The results show that this prediction is borne out in the data to a certain extent: a number of the subjects are clustering on $\lambda = 1$ while others cluster on $\lambda = \lambda_A^f$ over the sequence of questions. Thus

	All Choices			Only $\lambda < 1$		
	ETE	**ETO**	**WSRS**	**ETE**	**ETO**	**WSRS**
A's	0.314	0.249	6.86 (0.00)	0.182	0.124	5.58 (0.00)
B's	0.342	0.324	2.38 (0.01)	0.206	0.211	0.56 (0.29)
C's	0.256	0.206	4.83 (0.00)	0.191	0.151	4.15 (0.00)

Table 10: Average of $\lambda - \widehat{\lambda}_A^f$ and $\lambda - \lambda_A^f$ with z-statistic and associated p-value of a one sided Wilcoxon signed rank-sum test that $\lambda - \lambda_A^f$ is less than $\lambda - \widehat{\lambda}_A^f$.

we can take this behavior as an indication that the motivation behind the preferences of those subjects is similar to the ETO hypothesis.

The final piece of evidence is contained in examining the distributions of ϕ and $\widehat{\phi}$. These are the parameters representing a subjects preference for fairness as defined by $\phi = \theta_i/\kappa_i$ and $\widehat{\phi} = \hat{\theta}_i/\hat{\kappa}_i$.It is possible to compute a value for this parameter implied by a subject's choice of λ. Figures 7 and 8 show the distribution of ϕ and $\widehat{\phi}$ implied by the choices of the type A subjects for each question. In each cell of the Figures there is a dotted line at the critical value of ϕ or $\widehat{\phi}$, ξ or $\hat{\xi}$ respectively, that is a lower bound on the observable values. One note about the data displayed, the mass of all values of $\phi > 1$ or $\phi < -1$ has been truncated to those points.

In examining the distributions, there will appear a few violations of these lower bounds. Any negative values of ϕ or $\widehat{\phi}$ found in these graphs are derived from $\lambda's$ chosen at a point less than the most fair outcome associated with that notion of fairness. As such, any values of $\phi < 0$ or $\widehat{\phi} < 0$ indicate a choice that is difficult to justify based on that notion of fairness. For very large negative values, these represent choices of λ just slightly less than the most fair λ under that hypothesis and these may be reasonable due to our discretization of the choice space or just small errors. The choices that present a problem of interpretation are those that are negative and close to zero or positive and less than the

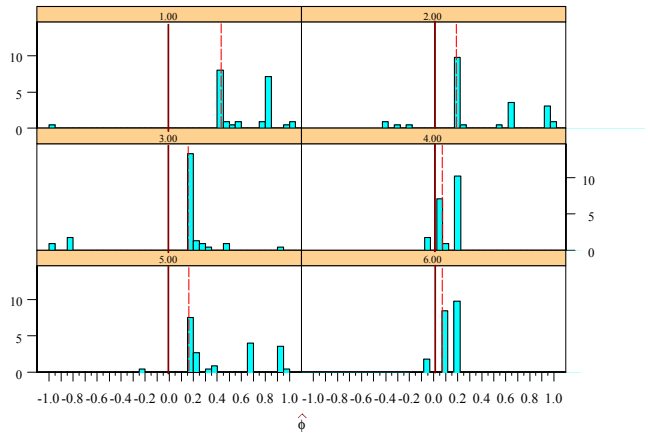

Figure 7: Histogram of the distribution of $\widehat{\phi}'s$ calculated for the type A subjects according to the equal treatment of equals hypothesis. Dotted line represents lower bound threshold for $\widehat{\phi}$, solid line denotes 0.

lower bound. Examination of these points indicate a choice of λ close to the lower endpoint of the chord. For some of the chords these points represent a lower level of fairness than the upper endpoint. Taken literally, this indicates that these individuals are malevolent in the sense that they are willing to make a sacrifice of their own probability of winning to make the overall allocation procedure less fair. For both notions, there are only a few such choices and these might be taken as extreme errors or misunderstandings.

One obvious characteristic of these distributions is that there is a much larger mass of observations in the distribution of $|\phi| > 1$ than for $\left|\widehat{\phi}\right| > 1$. This is just a reflection of the phenomenon reported earlier that found the choices of λ tend to cluster around λ^f more closely than they do $\hat{\lambda}^f$ as the measure of the intensity of fairness tends to approach infinity as the choice of λ approaches the most fair point for a particular fairness notion.

While there is no theoretical prediction made by either notion as to what these values should be, if we are proposing that either notion is an accurate reflection of the behavior

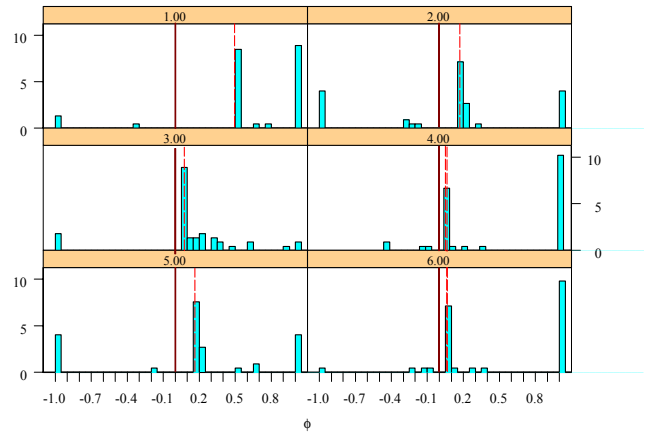

Figure 8: Histogram of the distribution of $\phi's$ calculated for the type A subjects according to the equal treatment of others hypothesis. Dotted line represents lower bound threshold for ϕ, solid line denotes 0.

of the subjects, it would certainly be advantageous if our measure of their preference for fairness remained fairly constant across questions. One method of testing to determine which of these preference notions best describes the behavior would be to determine if either hypothesis yields stable measures of the subjects preferences across questions. There is, however, an obvious problem with this test which has to do with the fact that the distributions of our observed values of ϕ and $\widehat{\phi}$ are truncated at different points for each question. A distribution test that does not correct for this problem will be biased. We are, however, not aware of a testing methodology to remedy this problem. With this caveat noted, we will proceed to determine if anything can be learned from these potentially biased tests. From examining the position of the truncation points, it appears that the most severe problem comes from the inclusion of question 1 as it possesses a much higher truncation point than for any other question while the truncation points for the other questions are clustered relatively close together. This suggests that if question

	Ques 1	Ques 2	Ques 3	Ques 4	Ques 5
Ques 2	3.389 (0.00)	–			
Ques 3	5.38 (0.00)	3.45 (0.00)	–		
Ques 4	5.35 (0.00)	4.51 (0.00)	2.62 (0.01)	–	
Ques 5	3.96 (0.00)	-0.24 (0.81)	-4.25 (0.00)	-5.56 (0.00)	–
Ques 6	5.36 (0.00)	4.52 (0.00)	2.52 (0.01)	0.05 (0.96)	5.56 (0.00)

Table 11: Results of Wilcoxon signed-rank tests of equality of distributions of comparing each question for Equal Treatment of Equals. Z-statistic (p-value).

1 were removed from consideration, this should reduce the bias significantly.

An initial test that one might run to examine this question would be a Kruskal-Wallis test to determine if the distributions of ϕ or $\widehat{\phi}$ remain constant over all six questions. Such a test in fact results in a strong rejection of stable distributions of the preference parameter measured under either ETE or ETO.[7] If the responses to question 1 are dropped from consideration, the parameters measured according to the ETE hypothesis still display no stability, but the hypothesis that the distributions of parameters measured on questions 2-5 according to the ETO hypothesis remain constant is not rejected at a 3% significance level.[8] In considering the result, one should keep in mind that due to the bias derived from the different truncation points, the p-values obtained are likley to be underestimated.

For a finer level of detail, we can test the equality of the distribution of preference parameters using Wilcoxon signed-rank tests between each pair of questions. The ETE results are contained in Table 11 and the ETO results in Table 12. In the ETE results, it is clear that most of the p-values are 0 with a few around .01 which means in general we can easily reject the hypothesis of equivalent distributions. There are only two sets of questions where we can not reject the null hypothesis, 2&5 and then 4&6. This is just

[7] The actual test statistics are 154.78 with p-value of 0.00 for the ETE hypothesis and for the ETO hypothesis, the test statistic is 42.61 with p-value of 0.00.

[8] The test statistics are 109.94 and p-value of 0.00 for the ETE hypothesis and 10.66 and a p-value of 0.031 for the ETO notion.

	Ques 1	Ques 2	Ques 3	Ques 4	Ques 5
Ques 2	4.30 (0.00)	–			
Ques 3	4.75 (0.00)	1.63 (0.10)	–		
Ques 4	2.61 (0.01)	-2.24 (0.25)	-3.36 (0.00)	–	
Ques 5	4.49 (0.00)	-0.13 (0.90)	-1.67 (0.09)	1.78 (0.08)	–
Ques 6	3.21 (0.00)	-2.23 (0.03)	-3.10 (0.00)	0.30 (0.77)	-1.25 (0.21)

Table 12: Results of Wilcoxon signed-rank tests of equality of distributions of f's comparing each question for Equal Treatment of Others. Z-statistic (p-value)

further confirmation of results found above that choices along these pairs of chords are indistinguishable as chords 2&5 are inverses and 4&6 are the same chord with different hypothetical prizes.

The ETO results have only a few p-values of 0 and all have either question 1 or 3 involved. We can reject the distribution of parameters found on question 1 as being the same as that found for any other question, and the parameters measured on questions 2 and 5 appear statistically different than those measured on question 3. In all other cases, we can fail to reject the null hypothesis by a reasonable margin. The reason for the differences between the choices on question 1 and all others can be attributed to the much higher truncation point. There does not appear to be any simple explanation for the results comparing questions 3 and 2 or 3 and 5.

We conclude that although there is no single strong definitive test available for distinguishing between these two theories, it does appear that the ETE notion of preferences finds little support in the data. The ETO notion, while not perfect, seems to be a closer characterization of the behavior of the subjects.

5. Conclusions

There is one important issue in the interpretation of our results that we have not yet considered. Due to the fact that player types were assigned randomly, some subject may

regard this assignment itself as an integral part of the allocation procedure that embodies the notion of equal treatment. If this is the case, then player A may feel justified taking full advantage of the situation in which he finds himself by choosing that allocation procedure that maximizes his probability of winning. This is an alternative explanation for the observed concentration of subjects of type A choosing the upper endpoint of the chord. This may also explain why some type C players indicate that this is a fair choice, thus explaining the puzzle as to why type C players regard the choice of the upper endpoint as fair. Perhaps the most striking experimental result then is that given this possible interpretation, still a significant number among subjects of type A, the "deciders," made choices involving a sacrifice of their own probability of winning to attain a fairer allocation procedure. This behavior lends support to the theory of self-interest seeking moral individuals of Karni and Safra (2000).

There appears little evidence in our data to support the equal treatment of equals preference notion. We do, however, find evidence supporting a slightly different specification of the ETO hypothesis can be seen by observing that the choices of λ across questions track the movements of λ^f very closely causing the distribution of choices to be bimodal with one of the modes typically being right on λ^f as predicted by the absolute value specification of the model. Further, the distribution of parameters measuring the intensity of preferences of the subjects show no tendency to remain stable when measured according to the ETE notion while there is some tendency for the parameters to remain stable when measured according to the ETO notion.

Our results on whether or not paying subjects based on their choices makes a difference are ambiguous. On question1, type A subjects choose in the exact same manner

whether they are being paid or not while on question 2, there is a significant difference. The assumption of symmetry, i.e., that type A treats types B and C impartially, is supported by the evidence.

The results further suggest the specific ways in which what subjects construe as "fair" choices are something that have to do with the social context in which they are choosing, and how such a context can be thought of as another layer on top of ones personal, selfish preferences. In particular, in the experiment, the context for type A players is that the payoff of the other subjects in the group is determined by their choices, and that seems to matter to them. The results show that this tendency towards fairness tends to carry over to all of the questions, even though they know that only the first question counts. The type B players, on the other hand, do not have to worry about this. Their answers are supposed to be what they would choose if they were the type A player, and they tend to act more selfishly. Type C player choices do not count either, but they were explicitly asked for what the *fair* allocation procedure is. The fact that the distribution of their answer is similar to that of the choices of the subject of type A may indicate a misunderstanding of what they are asked to do or that many of them believe that it is fair that the subject who has the power to decide to act selfishly. At any rate, the behavior of type C players and the study of what the subjects in such experiments consider to be "fair" allocation procedures warrants further examination.

APPENDIX

Experiment Instructions

The experiment is begun by recruiting a volunteer from the subjects to serve as a monitor for the experiment. Once a subjects has volunteered, the experimenter reads the

following message to the rest of the participants.

> Thank you for participating in today's experiment. In a few moments, I will ask you to turn to the computer screen in front of you and log-in to the system. Once you have done so, you will be lead through a series of help screens detailing the choice task you will be asked to engage in as well as the interface you will be using. Please practice with the interface so that you understand well how it works. You will be asked in this experiment to make 6 separate decisions concerning different lottery allocations. The first of these lotteries will be run according to the rules that will be detailed in the instructions. The next 5 will be hypothetical choices and will not be used to determine your earnings in this session. The lotteries will be run by the volunteer monitor using these two ten-sided dice to ensure that they are run fairly. The monitor will also observe and ensure that the proper amount of money is placed in each envelope.
> If there are any questions during the experiment, please raise your hand and I will come to assist you. Are there any questions at this point?
> Please log-in to the system and begin.

Once the subjects login to the computer system, they are presented with a series of help screens leading them through the experiment. The first is another introduction screen.

> You have volunteered to participate in an economic experiment on decision making. If you have any questions during the experiment please raise your hand and ask the proctor.
> In this experiment, you will be asked to make a series of 6 choices. For each of these choices, you will be asked to make a decision concerning the chances to win a prize for a group consisting of yourself as well as two other participants in the room. One of you will be designated as player A, one player B and one player C. With each new choice, the group you are in will change but your player type will remain the same throughout the experiment. Player A will be presented with a choice of how to allocate the chances of winning a prize between the three group members. Players B and C will both be making choices that will have no impact on who wins the prize. At the conclusion of the experiment, the probabilities chosen by the player A's for the first choice will be used to award the prize associated with that decision. Once the prizes have been assigned, everyone will be paid their show-up fee and winnings in cash.
> If at any point you have a question, please raise your hand and a proctor will help you. Please refrain from talking during the experiment and from looking at the screens of other participants.

After they press a button to continue, a version of the interface, Figure 9, is brought up with a box along the left hand side that contains text explaining how the experiment will work. The first block of text orients them to some of the content of the interface screen.

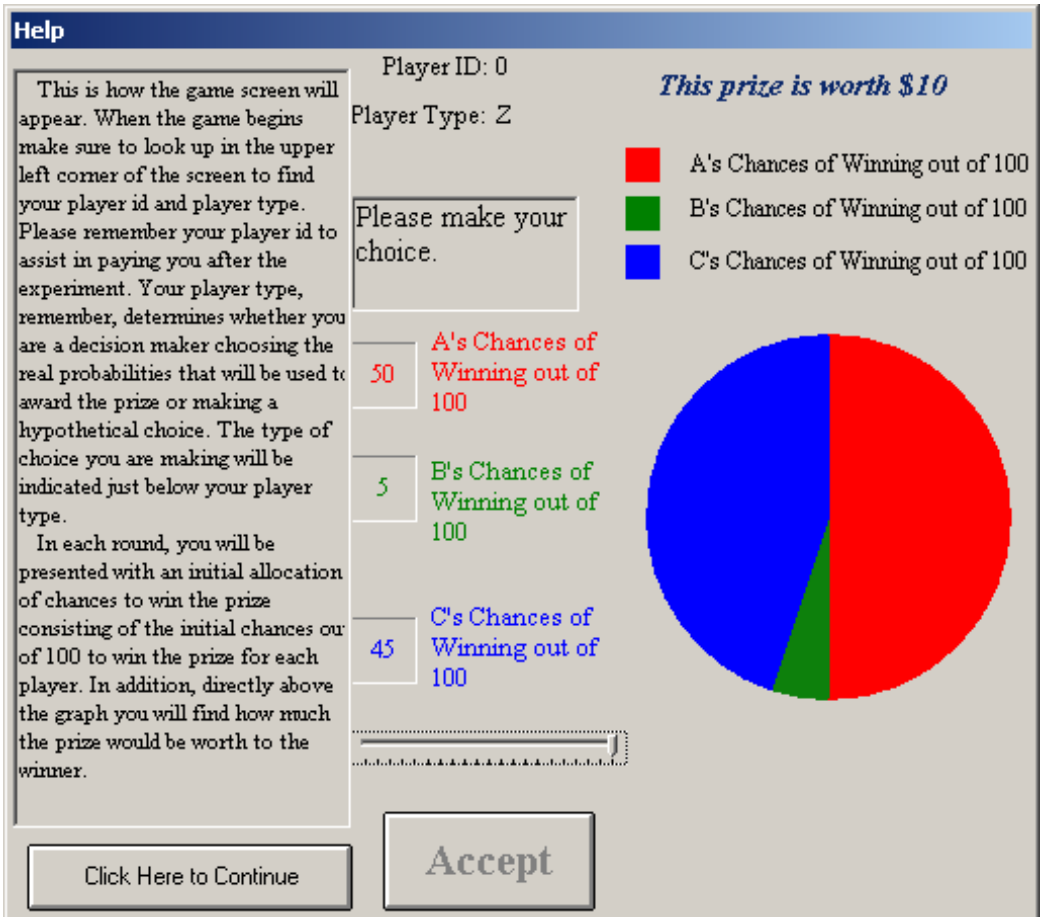

Figure 9:

This is how the game screen will appear. When the game begins make sure to look up in the upper left corner of the screen to find your player id and player type. Please remember your player id to assist in paying you after the experiment. Your player type, remember, determines whether you are a decision maker choosing the real probabilities that will be used to award the prize or making a hypothetical choice. The type of choice you are making will be indicated just below your player type.
In each round, you will be presented with an initial allocation of chances to win the prize consisting of the initial chances out of 100 to win the prize for each player. In addition, directly above the graph you will find how much the prize would be worth to the winner.

The next screen explains how players will make choices.

At the beginning of the experiment, you will be randomly assigned a player type

> that will remain constant throughout the experiment. In each period, however, the group you are in will change.
> If you are designated as a player of type A then you will be choosing how to allocate chances to win the prize in each period. At the beginning of each turn, you will be presented with an initial distribution. In this example player A has been allocated 50 chances to win, player B 5 and Player C 45. If you are player A, you will be able to use the slider bar at the bottom of the screen to change these probabilities.

The third screen explain in general terms what players B and C will be doing and has the players practice moving the slider bar. As the text indicates, players could not advance past this screen without moving the slider bar.

> If you are designated as a player B or C you will be asked to make a hypothetical choice. You will also do this by moving the slider bar. Your choices will have no impact on anyone's payoffs.
> Try moving the slider bar around now to see how it works. Notice that the graph on the right shows a pie chart representation of the possibility of each player winning. As you move the slider bar, the graph updates automatically as do the text boxes indicating the chances for each group member to win.
> Note: You must try moving the bar to continue.

The final screen explains how players submit their choices.

> Once you have made your choice for the allocation, click on the button labeled "Accept." You will be asked to confirm your choice before it is sent on to the server. When everyone's choices have been submitted, the groups will be reshuffled and you will move to the next choice.
> Try making a selection with the slider bar clicking on the "Accept" button now to see how it works. Clicking on the "Continue" button now will begin the game.

Once players advance past this screen, they enter into the actual experiment interface. Before all of the other subjects are finished with the instructions, all of the controls are greyed out and inactive and there is a dialog box on the screen asking the subjects to wait patiently. Once all subjects have finished the instructions, the dialog boxes disappear, the controls are enabled and the experiment begins.

References

Andreoni, James and John Miller (2000) "Giving According to GARP: An Experimental Test of the Rationality of Altruism," Working Paper.

Berg, Joyce, John Dickhaut and Kevin McCabe (1995) "Trust, Reciprocity, and Social History," *Games and Economic Behavior,* 10, 122-142.

Bolton, G. and A. Ockenfels (2000) "ERC: A Theory of Equity, Reciprocity and Competition," *American Economic Review,* v90, n1: 166-93

Bolton, G. and A. Ockenfels (1998) "Strategy and Equity: An ERC-analysis of the Güth-van Damme Game," *Journal of Mathematical Psychology*, Vol 42, 215-226.

Camerer, Colin F. (1997) "Progress in Behavioral Game Theory," *J. of Econon. Perspectives,* 11, 167-188.

Elster, Jon (1998) "Emotions and Economic Theory," *J. of Econ. Literature,* 36; 47-74.

Engelmann, Dirk and Martin Strobel. (2001) "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments," mimeo

Fehr, E. and K. Schmidt, (1999) "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, Vol 114, 817-868.

Güth, W. and E. van Damme (1998) "Information, Strategic Behavior and Fairness in Ultimatum Bargaining: An Experimental Study," *Journal of Mathemtical Psychology*, Vol. 42, 227-247.

Kagel, J. and K. Wolfe (1999) "Testing Between Alternative Models of Fairness: A New Three Person Ultimatum Game," Working Paper

Huber, P.J. (1967) "The behavior of maximum likelihood estimates under non-standard conditions." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley, CA: University of California Press, 1, 221-233.

Hume, David (1740) *Treatise on Human Nature.* London, J. M. Dent and Sons, 1939.

Karni, Edi and Zvi Safra (2000) "Individual Sense of Justice: A Utility Representation,"

——————————— (2000a) "Intensity of the Sense of Fairness: Measurement and Behavioral Characterization,"

Loewenstein, George (2000) "Emotions in Economic Theory and Economic Behavior," *The American Economic Review; Papers and Proceedings. 426-432.*

Rawls, John (1963) “The Sense of Justice,” *Philosophical Review,* 72, 281-305.

————————- (1971) *A Theory of Justice.* Cambridge, Harvard University Press.

Romer, Paul M. (2000) “Thinking and Feeling,” *The American Economic Review; Papers and Proceedings. 439-443.*

Smith, Adam (1759) *The Theory of Moral Sentiments.* New Edition, D. D. Raphael and A. L. Macfie (eds.) Oxford, Oxford University Press, 1976.

Smith, Vernon L.(1976) ”Experimental Economics: Induced Value Theory,” *American Economic Review,* 66 (2): 274-79.

Sopher, Barry and Mattison Narramore (2000) “Stochastic Choice in Decision Making Under Risk: An Experimental Study,” *Theory and Decision*, 48, 323-350.

White, Halbert (1980) ”A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica*, 50: 1-25.