Exp Econ (2008) 11: 174–189
DOI 10.1007/s10683-007-9165-1

Individual sense of fairness: an experimental study

Edi Karni · Tim Salmon · Barry Sopher

Received: 16 March 2006 / Accepted: 13 February 2007 /
Revised: 6 February 2007 / Published online: 14 July 2007

Abstract Many prior studies have identified that subjects in experiments demonstrate preferences for fair allocations. We present an experimental study designed to test whether a similar concern for fairness manifests itself when the decision maker is choosing among differing probabilistic allocation mechanisms that will all generate an *ex post* unfair allocation by assigning an indivisible prize to one individual. This investigation is inspired by Karni and Safra (Econometrica, 70, 263–284, 2002) in which a structure for preferences for fairness in such an environment was developed. Here we use this model to design experiments that allow us to test for the presence of concern for fairness in individual choice behavior and examine some factors that may affect the intensity of the concern for fairness.

Keywords Dictator game · Preferences for fairness · Choise under uncertainty

JEL Classification C91 · D63

Electronic supplementary material The online version of this article (http://dx.doi.org/10.1007/s10683-007-9165-1) contains supplementary material, which is available to authorized users.

E. Karni (✉)
Department of Economics, Mergenthaler 469, Johns Hopkins University, 3400 N. Charles St.,
Baltimore, MD 21218, USA
e-mail: karni@jhu.edu

T. Salmon
Department of Economics, Florida State University, Tallahassee, FL 32306-2180, USA
e-mail: tsalmon@fsu.edu

B. Sopher
Department of Economics, Rutgers, The State University, 75 Hamilton St., NJ Hall Room 202,
New Brunswick, NJ 08901-1248, USA
e-mail: sopher@economics.rutgers.edu

1 Introduction

By and large, neoclassical economics is founded on a narrow notion of self-interest seeking behavior, where self-interest is defined in terms of *material well-being*. This 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] There is, however, 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 showing that subjects do not always make choices consistent with narrow definitions of pure self-interest.[2]

In this paper we explore, via an experiment, 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 involving random selection of a winner of a predetermined prize, we look for evidence of willingness to sacrifice one's own chance of winning to attain what is perceived to be a fairer allocation procedure. Our subjects participate in a three-person dictator game in which one of the three players chooses a lottery that is used to determine who, among the three, wins a $15 prize. This work is inspired by two recent papers of Karni and Safra (2002a, 2002b). The first paper presents an axiomatic model of choice among random allocation procedures of individual motivated, in part, by concern for fairness. The second paper introduces measures of the intensity of individual sense of fairness and derives their behavioral characterizations. Our experimental design is based on the analytical framework of Karni and Safra (2002a) and is intended to test their contention, that inherent sense of fairness is manifested in individual choice among random allocation procedures.

Some studies of three person dictator games or variants of ultimatum games (Güth and van Damme 1998, Bolton and Ockenfels 1998, 2000 and Kagel and Wolfe 2001) conclude that the proposer tends to ensure rough equality between the two recipients. Other studies of three person dictator games (see Charness and Rabin 2002 and Engelmann and Strobel 2004) attempt to distinguish between various models of fairness.

Our method is similar to that of Andreoni and Miller (2000), whose 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. In a related series of papers Fisman et al. (2005a, 2005b, 2005c) use a similar approach for testing issues involved in preferences for fairness. These contain graphical representations of 2 and 3 person dictator games in which the dictator choices are restricted to a budget sets that involve varying the price for trading off welfare for the dictator and the recipient.

[1] See Hume (1740), Smith (1759), Rawls (1963, 1971) for philosophical discussions.

[2] A comprehensive list of such papers is too long to include, but a good introduction to this literature can be found in Chap. 2 of Camerer (2003).

Our design has a similar interpretation: the line segment, in the probability simplex, is the equivalent of a budget constraint and the dictator's choice represents the point in that budget constraint that intersects his "highest" indifference curve. However, unlike both Andreoni and Miller and Fisman et al. who are mainly interested in determining the consistency of choices in deterministic environments, our main interest is the manifestation of a concern for procedural fairness as proposed in Karni and Safra (2002a). This requires a new experimental design in which the choices involve uncertainty in a manner that was not studied previously. Many studies provide evidence suggesting the existence of preferences for fairness in deterministic allocations. The current study examines the presence of a concern for fairness in individual choice among random allocation procedures. We note that, if one's sole concern is the fairness of the outcome, then any procedure that assigns a prize to one among equally deserving individuals is equally unfair. Even if a person is willing to compensate another person to make the ex-post allocation fairer, it does not necessarily follow that the same person would be willing to reduce his chance of winning a prize to improve the chances of others. Our study focuses on this issue.

Bolton et al. (2005) also studied individual preferences over allocation mechanisms using experimental methods. Subjects were asked to choose between different discrete procedures for dividing an amount of money, including an ultimatum game in which the offer is determined by lottery and games that allow the proposer to choose between having a lottery make an initial offer or making the offer themselves. The focus of this work is on the receiver's view of their treatment at the hands of the proposer in terms of "fairness," or more properly the "acceptability," of different offers depending on the mechanism through which the offer is made. Our interest is the assessment of the preference structure of the individual making the offer. Consequently, we avoided the use of a setting which involves interactive decision making. These are complementary lines of research aimed at different aspects of the broad question of how people view the fairness of different procedures.

The rest of the paper is organized as follows: In Sect. 2, we will provide a brief description of the theoretical environment from Karni and Safra (2002a) and describe the experiments designed to test the theory. In Sect. 3, we present and analyze the findings. The main conclusions and issues raised by this work are summarized in Sect. 4.

2 Theory and the design of the experiments

2.1 Theory

To set the stage, we review briefly those elements of the theory of Karni and Safra that are relevant for the current study. We focus on aspects of the model that underlie the experimental design.

Let $N = \{1, \ldots, n\}$, $n > 2$, be a set of individuals who must decide on a procedure by which to allocate, among themselves, one unit of an indivisible good. Because only one individual is awarded the good, the ex-post allocation is necessarily unfair. The issue, therefore, is what allocation procedure may be implemented to attain a higher level of fairness ex-ante. Karni and Safra (2002a) restrict attention to procedures that allocate the good by lot. Formally, let e^i, the unit vector in $\mathbb{R}^n$, denote the *ex-post* allocation that assigns the good to individual i. 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*, or allocations by lot.

An individual, in this model, is characterized by two binary relations, $\succcurlyeq$ and $\succcurlyeq_F$, on P. The preference relation $\succcurlyeq$ represents his actual choice behavior and the relation $\succcurlyeq_F$ represents his conception of fairness. The preference relation $\succcurlyeq$ has the usual interpretation, namely, for any pair of allocation procedures p and q in P, $p \succcurlyeq q$ means that, if he were to choose between p and q, the individual would either choose p, or be indifferent between the two. The fairness relation, $\succcurlyeq_F$, has the interpretation of "fairer than," that is, $p \succcurlyeq_F q$ means that the allocation procedure p is regarded by the individual as being at least as fair as the allocation procedure q. The notion of fairness and how intense the sentiment for fairness is, are subjective, intrinsic and, together with concern for self-interest, governs the individual's behavior.

Taking the preference and the fairness relations as primitives, Karni and Safra (2002a) derive a third binary relation, $\succcurlyeq_S$ representing the self-interest motive implicit in the individual 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 $\kappa : P \to \mathbb{R}$ representing the relation $\succcurlyeq_S$, a strictly quasi-concave function $\sigma : P \to \mathbb{R}$ representing the fairness relation $\succcurlyeq_F$, and a utility function V representing the preference relation $\succcurlyeq$ as a function of its self-interest and fairness components, i.e., for all allocation procedures, $p, q \in P$,

$$p \succcurlyeq q \Leftrightarrow V((\kappa \cdot p, \sigma(p)) \geq V((\kappa \cdot q, \sigma(q)).$$

In addition, Karni and Safra (2002a) examine the case in which the function V is additively separable in the self-interest and fairness components. Formally,

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

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 must select the allocation from a predetermined set represented by a line segment in the probability simplex allowing him to trade off his own chance of winning to improve the overall fairness of the allocation procedure. We will be testing whether the preference structures display the strict quasi-concavity hypothesized in Karni and Safra (2002a, 2002b).

2.2 Experimental design

The design of these experiments is a three person dictator game, which is executed using a modified version of the interface used in Sopher and Narramore (2000). 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 were then led through an interactive help program to make sure that they understood the interface and rules of the experiment. Upon completing the instructions sequence, each subject was 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 subject A in each group, whose behavior is the main concern of this study, was asked to choose the allocation of the probabilities to the subjects in the group to be used in the actual lottery for a $15 prize. More specifically, the subjects of type A are asked to design a lottery $p = (p_{\mathrm{A}}, p_{\mathrm{B}}, p_{\mathrm{C}})$, where $p_i \geq 0$, $i = \mathrm{A, B, C}$ and $\sum_{i=1}^{3} p_i = 1$, to be used to select the winner of the $15. In this context, p_i is the probability that subject i wins the prize. Type A's instruction was "Please choose the allocation of chances to be used in deciding who among A, B, and C wins the prize."

The subjects B and C in each group were asked to perform similar tasks, but their choices did not affect their own or the other players' payoffs. In particular, subjects of type B were instructed to: "Please select the allocation that you would choose if you were the decision maker, subject A," and subjects of type C were instructed to "Please select the allocation that you believe is fair." The purpose of asking subjects of type B and C to perform a choice task was to prevent them from identifying A by observing some subjects making a choice and others not. As a secondary consideration, it is interesting to examine the preferences expressed by the other subjects in a hypothetical context, and to elicit their views on what the fair allocation procedure is. Note that, it was important, in our design, that A's choice is the sole determinant of the final payoff of B and C. Thus these subjects were not rewarded for their performance. The nature of A's choice was common knowledge as was the fact that B and C players would be responding to a question with a similar structure.[3]

The choice set, corresponding to this design, is a 2-dimensional simplex depicted in Fig. 1. The top vertex of the triangle represents the allocation procedure according to which the subject A is the sure winner. Similarly, the lower left and right vertices are the allocation procedures that making subjects B and C, respectively, the sure

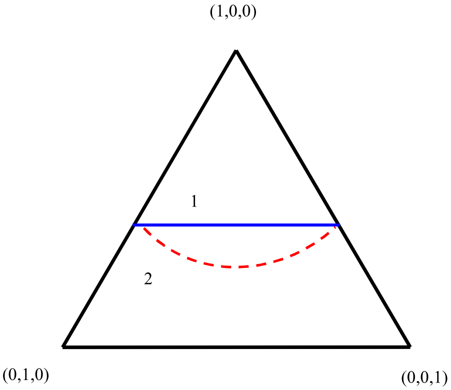


Fig. 1 Characterization of indifference curves. *Line 1* characterizes a person with no preference for fairness. *Line 2* characterizes a person who exhibits a preference for fairness

[3] A complete record of the help screens that the subjects were led through to explain the experiment is available from the authors in a supplementary appendix.

winners. Supposing that the three subjects perceived themselves to be equally deserving of the prize, the intensity of their sense of fairness can be represented, in this context, by the curvature in their indifference curves.[4] At one extreme, if the subject A exhibits no sense of fairness, then his indifference curves will be straight lines, such as line 1, along which p_A is constant. However, if he is concerned about the fairness of the allocation procedure, and regards the two other subjects in the group as equally deserving of the prize then, assuming that his preferences are continuous, the indifference curves may be convex, as shown by line 2 indicating strict preference for fairness.[5] This indifference curve depicts a willingness to sacrifice one's own probability of winning to attain a fairer overall allocation procedure.

In the experiment, a subject is presented with a line segment, 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 of winning. If, however, the subject is concerned about the probabilities with which the other players might win then the optimal lottery may be represented by a point in the interior of the line segment.

The subjects were asked to make a total of six choices along 5 different line segments. These are depicted in Fig. 2 and the lotteries defining the endpoints of the line segments are shown in Table 1. The line segments chosen possessed some specifically designed similarities to allow the investigation of specific issues to 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 (that is, a subject who was assigned type A at the outset remains type A for all six trials).

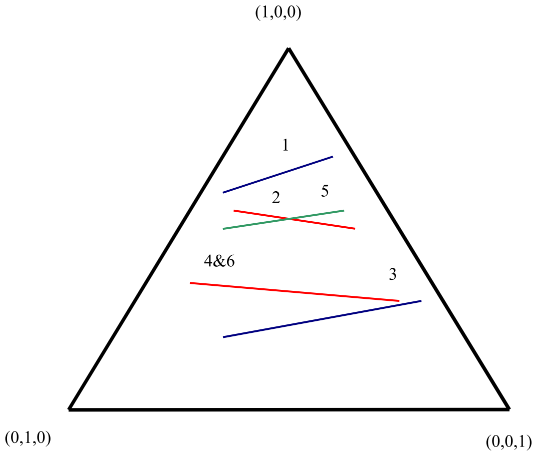


Fig. 2 Graphical representation of the line segments used in the experiment

[4] See Karni and Safra (2002b) for a detailed analysis.

[5] The continuity of the preference relation rules out lexicographic orderings under which concerns for fairness may be a dominated concern. In general, distinct points on an indifference curve represent tradeoff between self-interest and fairness. The strict curvature of the indifference curve depicted in Fig. 1 is implied if the upper counter sets of the fairness relation are strictly concave. The theory does not rule out that the indifference curves consist of two line segments, provided that the "better than" is strictly convex.

Table 1 Lotteries defining the endpoints of the line segments used in the experiments

	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

The use of multiple sequential choices raises the possibility that subjects could engage in behavior based upon compounding the lotteries across choices. This could have lead the A subject to think that he was being fair by staying at the starting endpoint of the line segment in all of the choices as this might equalize the chance of winning for subjects B and C across the experiment. Consequently, even if A did not move from the endpoint, it would not have been possible to conclude that these choices were not motivated, in part, by a concern for fairness. To overcome this problem, only the first lottery was used to generate actual payoffs. The choices made by the type A subjects 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.

Because only one line segment was used to select 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 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 test the degree to which subjects display a stronger preference for fairness when the decision is hypothetical versus when it is real.

To make the choice easy to understand, 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 integer values and as colored slices of a pie. Subjects could use a slider bar to move along the line segment 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. The final choice of a subject 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 type A subject chose the point that maximized his probability of winning while a choice of $\lambda = 0$ indicates that he chose an allocation procedure that minimized his chance of winning. Note that lower values of λ (down to some point) correspond to greater equality of the probability of winning assigned to subjects of types B and C.

The subjects used in these experiments were drawn from two separate populations. 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 135 subjects participated in these experiments, with 69 from CIT and 66 from PCC. Each session included subjects from only one population 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 the PCC subjects the show-up fee was $10.[6] The sessions for these experiments lasted from 20 minutes up to 40 minutes. Most sessions lasted between 20 and 30 minutes.

3 Results

3.1 Methods

The experimental results are choices of probability mixtures $(\lambda, (1-\lambda))$ of two lotteries, where, for each question, λ denotes the weight on the lottery that gives A the distribution that first-order stochastically dominates all the other feasible distributions in his choice set. Underlying these choices, we hypothesize, 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. Our ordered probit estimation procedure amounts to estimating a noisy version of the intensity of the sense of fairness as developed in Karni and Safra (2002b).

In the estimated model, we include indicators for each distinct question in the experiment to capture any variation in choice behavior due to the differences in budget constraints. Other regressors include: (i) a dummy variable, MALE, equal to 1 if the subject was male and 0 for female; (ii) a dummy variable, PAY, equal to 1 if the question will result in payment, and a variable PAY*Q2, indicating if the paid question was Q2; (iii) a dummy variable, ORDER, equal to 1 if Q1 was first and 0 if Q2 was first. Although we have not provided a formal account of the ordered probit model, it is nonetheless the case that the signs of the coefficient (given that the underlying regressor is always positive) indicate more selfish choices when positive, less selfish choices when negative, compared to the baseline choice of Question 1, not paid.

We estimate a random effects ordered probit model. The random effects specification means that we treat the error, ε_i, as being composed of two parts, an individual-specific component, which is the same for every observation on an individual, and an idiosyncratic component which varies over different observations on an individual.[7] Since choices are ordered along each line segment in the simplex, we can treat each mixture choice with positive mass in the distribution of choices as a discrete choice. In fact, however we have reduced the full set of discrete choices observed into 9 categories, as shown in Table 2. We have done this because the random effects estimation cannot be performed if the cell counts in a category are too small. Henceforth, we may refer to these new choice categories as the λ choice of the subject, though it

[6] 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.

[7] We use the REOPROB procedure, an ADO routine in Stata written by Guillaume Frechette.

Table 2 Bins used for choices of λ in the ordered probit regression with number of choices in each bin by subject type

Choice category	Type A	Type B	Type C
0 ($0 \leq \lambda \leq .15$)	3	6	3
1 ($.15 < \lambda \leq .25$)	3	7	9
2 ($.25 < \lambda \leq .35$)	8	13	11
3 ($.35 < \lambda \leq .45$)	6	0	12
4 ($.45 < \lambda \leq .65$)	118	73	145
5 ($.65 < \lambda \leq .75$)	7	6	6
6 ($.75 < \lambda \leq .85$)	6	6	5
7 ($.85 < \lambda \leq .95$)	23	10	8
8 ($.95 < \lambda \leq 1$)	96	140	71

Table 3 Ordered probit results for choice mixture chosen

Indep. variable	Choice	
	Coefficient	p-Value
MALE	0.92	< 0.01
CIT	1.04	< 0.01
PAY	0.37	0.40
PAY*Q2	0.54	0.42
ORDER	−0.89	0.01
Q2	−0.52	0.43
Q3	−0.07	0.35
Q4	−0.44	0.36
Q5	0.19	0.36
Q6	−0.42	0.36
Groups (OBS)	45(6)	
LL	−306.98	
p-Value	.00	

should be kept in mind that each of these choice categories correspond to a range of choices in the actual experiment.

Table 3 contains results of the random effects ordered probit regression model which help us summarize efficiently the within- and between-question differences. We have only presented the regression results for type A subjects because theirs was the only choice that was incentivized. We briefly describe the choice behavior of the B's and C's below but will subject them to no analysis. The dependent variable is CHOICE, the mixture category corresponding to the chosen λ, while the regressors are as explained above. The coefficients and corresponding p-values for the z-statistics are reported for each variable. The table also reports Groups (OBS) (the number of groups for the random effects, i.e. subjects, and the number of observations per subject), LL (log-likelihood for the estimated model), and the p-Value, the probability associated with the model chi-squared test for the regression.

The regression for A subjects has significant coefficient estimates on MALE, CIT, and ORDER. The positive sign on the first two indicate that males and CIT students,

Table 4 Marginal effects on probability of choice for modal choices. * indicates significance at the 1% level

Player type	Type A	
Category chosen	Cat. 4	Cat. 8
MALE	−33%*	+30%*
CIT	−37%*	+32%*
PAY	−14%	+12%
PAY*Q2	−31%	+26%
ORDER	+28%*	−21%*
% of choices	44%	36%

on average, made significantly more selfish choices. The negative signs on the ORDER variable indicate that the group of subjects for whom Question 2 was the paid question were significantly less selfish (on all questions, on average). The regression includes estimated "cut-points," which are essentially constant terms for each of the discrete choice categories (and correspond to the threshold values discussed earlier). We do not report these, except to note that a majority of them are significant.

While the signs of the coefficients indicate the general nature of the shift in behavior, towards higher probabilities of more selfish choices, 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. There are 45 discrete choice values in the subject choices, and we have reduced these to 9 distinct categories. We compute marginal effects for the two modal choices: category 4, corresponding to choices between .45 and .65 (about 41% of all choices) and category 8, corresponding to choices between .95 and 1 (about 39% of all choices), which together account for 80% of all choices. Typically, we find (for the significant coefficients) that a positive sign on the coefficient corresponds to an increase in the probability that category 8 is chosen, and a decrease in the probability that category 4 is chosen, while a negative sign indicates the reverse.

Table 4 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% level or better by an asterisk. In general, the marginal effects are computed with the relevant dummy variable equal to one vs. zero, and all other variables at their sample means. The one exception is the marginal effect for PAY*Q2, where we set the PAY variable set to one or zero along with the PAY*Q2 variable. Thus, for this variable only, the marginal effect is very specifically the effect of pay on choosing for question 2.

An example of how to read these results is that male type A subjects are 30% more likely to choose category 8 (and 33% less likely to choose category 4) than female subjects. Similarly, CIT students are 32% more likely to choose category 8 (and 37% less likely to choose category 4) than are Pasadena City College students. In general, the significant effects are large, which is not surprising, given the concentration of choices in the two categories in question.

3.2 Further analysis

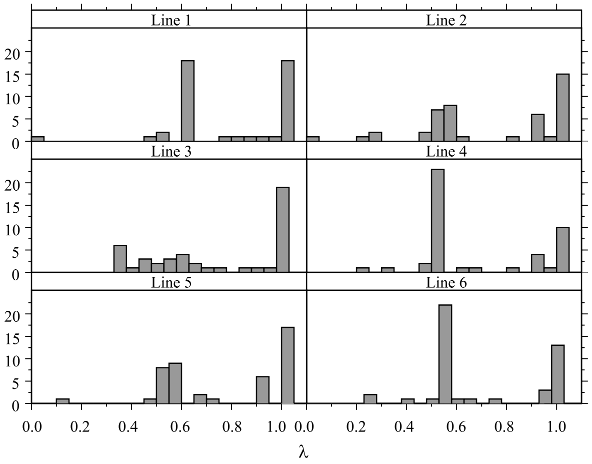
Next we take a closer look at the results from our experiment by presenting histograms of the choice behavior in order to convey the general structure of the data. We will then proceed through the main questions of interest using the results from the previous regressions and additional tests in order to establish a clear answer to each question.

3.2.1 Comparison of choices between subject types

Figure 3 contains a histogram of the choices made by the A players for all questions. We have also included Figs. 4 and 5 which contain the same for both players of type B and C.[8] We present the latter two just for completeness in presenting the data. We will include some descriptive discussion of these data but due to the uncertainty involved in understanding the choices by B and C players, we will subject them to no formal analysis.

On all of the questions, the distribution of A and C choices appear to be remarkably similar. For these subjects there is a bimodal distribution to the choices with one mode

Fig. 3 Histograms of choices for players of type A over all line segments



[8] Before proceeding, one unanticipated, feature of the data must be noted. Subjects B were asked what they would choose if they were subjects A, the deciders. In designing their role we intended them to suppose themselves in the position of A and make their choices accordingly. It seems, however, that some B 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 B and getting the chances of winning for a B subject from any given allocation. On line segments choices that slope down to the left (1, 3 and 5), some B subjects (12% of them) chose the lowest point on the line segment. Such choices maximize their chances of winning the prize while minimizing the chances of subject A to win the prize. This behavior makes no sense from the point of view of subject A, 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 A subjects 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 $\lambda = 1$ which we have done.

Fig. 4 Histograms of choices for players of type B over all line segments

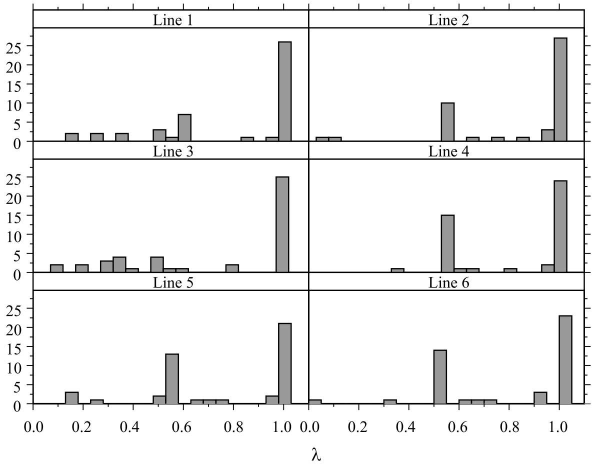
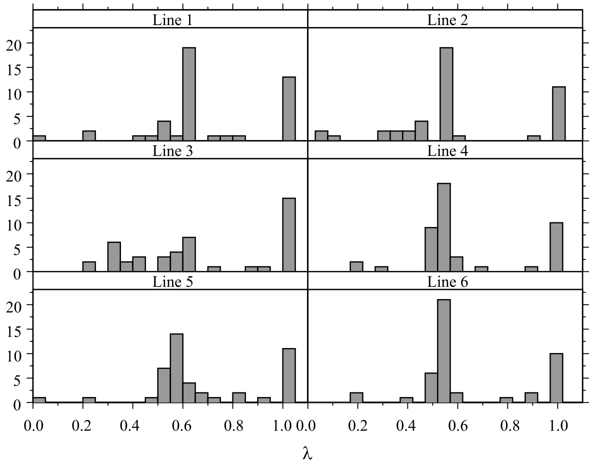


Fig. 5 Histograms of choices for players of type C over all line segments



at $\lambda = 1$ and this has the natural interpretation of being the most selfish choice for the A subject. The second mode for each question occurs at a λ that comes closest to equalizing the winning probabilities for the B and C players on that line segment. We refer to this point as the LETO for a question as it is the most fair λ associated with a notion of fairness involving Equal Treatment of Others.[9] The distributions for B subjects are also, typically, bimodal, but with a larger mode at 1 than is the case for A or C subjects. Note that these are distributions of the original choices, not the categories used in the regression analysis.

As already mentioned, our main concern is with the behavior of subjects of type A whose choices affect the ultimate payoffs. Due to the fact that choices of λ were restricted to the set [0, 1], all interior λ choices can be interpreted as near optimal

[9] The values of the LETO for each question are as follows: Q1—0.60, Q2—0.54, Q3—0.33, Q4&6—0.52 and Q5—0.54.

choices for the subject.[10] The large number of choices at 1 may well be the result of a censoring effect of the possible choices as the constraint may be binding for these subjects. To grasp the censoring effect recall that ϕ_i is a measure of the intensity of the sense of fairness. It may be assumed to be distributed over the half-open interval $[0, \infty)$. Given a line segment $(\bar{p}, \bar{q})$, let $\lambda^*(\bar{p}, \bar{q}; \phi_i)$ be the optimal choice of subject i who is assigned the role of A. Clearly, $\lambda^*(\bar{p}, \bar{q}; \cdot)$ is a monotonic decreasing function of ϕ_i, that is, the more intense is the individual sense of fairness, the more he is willing to sacrifice his own chance of winning to attain a fairer allocation procedure.[11] Hence the distribution of ϕ_i induces a distribution on $\lambda(\bar{p}, \bar{q}; \phi_i)$. However, the actual range of $\lambda(\bar{p}, \bar{q}; \phi_i)$ is $[\lambda_i^f(\bar{p}, \bar{q}), 1]$, where $\lambda_i^f(\bar{p}, \bar{q})$ denotes individual i's fairest allocation procedure. Let $\bar{\phi}_i$ denote the value that satisfies $\lambda(\bar{p}, \bar{q}; \bar{\phi}_i) = 1$. Then, the effect of censoring on the induced distribution of $\lambda(\bar{p}, \bar{q};)$ is that it tends to have a concentration at 1. Specifically, $\Pr\{\lambda(\bar{p}, \bar{q}; \bar{\phi}_i) = 1\} = \Pr\{\phi_i \leq \bar{\phi}_i\} = F(\bar{\phi}_i)$, where F denotes the cumulative distribution function of ϕ_i. It was the likely presence of this type of censoring, as well as the milder censoring due to the discrete choice set that subjects faced, that led us to use the ordered model to estimate the various effects in the previous section.

Perhaps the most important finding, regarding the choices of the A subjects, 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. This confirms the hypothesis in Karni and Safra (2002a) that indifference curves in this environment may be curved instead of horizontal as detailed in Fig. 1. Due to the nature of the choice task, it was not possible to conduct enough incented choices to construct a map of the space of indifference curves, but this result is enough to show that curvature exists.

3.2.2 Paid versus unpaid questions

In view of the fact that five of the six questions the subjects answered generated no payoffs, it is natural to ask what, if any, effect this had on the answers. A standard hypothesis from "induced value theory" (Smith 1976) is that subjects behave more selfishly when the choice has real consequences. 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 3. The coefficients on both the PAY and PAY*Q2 variables are insignificant for A subjects. The interpretation of these results is that the subjects who made choices on Q1 and Q2 when they were paid did not choose in a manner that is significantly different from those subjects who were not paid based on those decisions.

There might appear to be a complication to this issue in that the ORDER variable is negative and significant. This means that those subjects who saw Q1 first were, on

[10] Only "near" optimal because the choice set of λ's was limited to discrete choices.

[11] For a formal proof of this assertion see Karni and Safra (2002b).

average, less selfish than those who saw Q2 first. Note, however, that this has nothing to do with whether Q1 was first and paid or Q2 was first and paid, but rather reflects an average difference between the subsamples. This is because the estimated effect takes into account response on all questions, not just the paid questions. One possible explanation for this ORDER effect is that in question 1, the A player begins in a greater position of "wealth" relative to question 2 in that the probability of winning for the A player on question 1 is always greater than on question 2. It is possible that not only were people when in such a position more willing to be generous towards the other players but that this generosity carried forward to other questions as something of an imprinting effect.

Overall, the evidence indicates that there was no difference in behavior between our hypothetical and paid questions for type A players. This allows us to examine the choices we see on all of the hypothetical line segments without having to construct a correction for any hypothetical bias. Again, we chose this approach of paying only on the first choice and not on all or a randomly selected choice as a means of eliminating any lack of independence issues across questions. These results tell us that by doing so we did not create any other problems from the hypothetical nature of most of the questions.[12]

3.2.3 Symmetry

Once we allow departures from the standard self interest model, it is important to verify that the general structure of preference theory still holds leading to consistency of choices. In our experiment consistency in the form of symmetry requires that behavior on questions 2 and 5 should be similar as these line segments are identical except the positions of the B and C subjects have been reversed. Because B and C subjects have identical status in the game, the A subjects should see no reason to treat one differently from the other. Thus if the A subjects do treat B and C symmetrically, their choices on the two line segments should be similar.

The estimated coefficients in the ordered probit model, shown in Table 3, are insignificant. The histograms in Fig. 3 look remarkably similar, which tends to reinforce our conclusion that there is no significant difference here. A Wilcoxon signed rank sum test and a Kolmogorov–Smirnov test for the differences in the distribution of λ's across questions 2 and 5 for type A subjects results in p-values of 0.25 and 0.995 respectively. These tests were performed on the raw choice data, not based on the bins used in the estimation, and the results confirm the indication from the histograms and the regression results that the A players do treat the B and C players symmetrically.

4 Conclusions

The main purpose of this study was to test for the existence of preferences for fairness over random allocation procedures, using experimental methods. This is differ-

[12] We also tested for the effect of the size of the prize by questions 4 and 6 that are almost identical questions, with the only difference being the size of the hypothetical prize. On question 4, the prize was $15 while on question 6 the prize was $45. The results show no statistically significant difference in the choices on these two questions that can be attributed to the size of the prize. One possible explanation for this is the small difference in the size of the prize combined with the fact that the choice is hypothetical.

ent from most of the literature dealing with individual sense of fairness, in that our design tests the subjects' response to fairness of the procedure rather than that of the ultimate allocation. Our results show that, in these situations, a substantial proportion of subjects are willing to sacrifice their own probability of winning to effect a fairer overall allocation procedure.

While results suggesting that subjects' conduct is governed, in part, by a sense of fairness are by no means new, ours do possess some novel characteristics. Compared to standard two-person dictator game results, such as Forsythe et al. (1994), in which subjects give up approximately 20% of the certain pot of money to make the allocation more fair, our subjects in contrast look relatively selfish. In this experiment, subjects are required to give up relatively little, in expected value terms, to make the overall allocation procedure substantially fairer, yet a significant number of our subjects display no willingness to do so. This indicates that preferences for fairness, in this context, may not be as strong as in environments in which the decision maker is dividing certain amounts of money. One possible explanation for this is that the two recipients in this game never observe the probabilities chosen by the decision maker, only the outcome. Thus whether the decision maker is fair or not can not be ascertained by the recipients. This separation between choice and outcome may induce decision makers to be less fair.

There are two mechanisms that might deliver this outcome. One is contained in the results of Hoffman et al. (1994), in which the authors show that, by increasing the social distance between the decision maker and the recipient, the offers in the dictator game went down. By not showing the recipients the choice of A, we generated what amounts to substantial social distance in the form of cover on the part of A. Specifically, A could win the prize whether or not he behaved selfishly. The participants do not know. Alternatively, the results in Dana et al. (2004) suggest that if a decision maker can make someone else at least partially responsible for the outcome (in this case, the randomization process or the experimenter) then he acts substantially more selfishly or that behavior consistent with preferences for fairness "decreases substantially when the connection between choices and outcomes is obfuscated." In our experiment, the connection between choices and outcomes is substantially obfuscated through the use of the probabilistic allocation. Since the apparent strength for preferences for fairness decreases, it suggests that the decision maker is not necessarily concerned with fairness in itself, but rather is concerned with not appearing unfair in the minds of others. Our results are consistent with those of Dana et al. in this regard.

There is one important issue in the interpretation of our results, mentioned in Sect. 3.2, that merits further discussion. The fact that the roles of the subjects were assigned randomly, might make them regard the assignment procedure itself as an integral part of the allocation procedure, embodying the notion of equal treatment. In this case, subject A may feel justified in 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 A subjects choosing the upper endpoint of the line segment. This may also explain why some C subjects, indicate that this is a fair choice, thus explaining the puzzle as to why many C subjects regard the choice of the upper endpoint as fair. Perhaps the most striking result then is that, given this possible interpretation, still a significant number among the A subjects, the "deciders," chose to sacrifice some 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 (2002a).

Acknowledgements Useful comments of two anonymous referees and the editor are gratefully acknowledged. We would also like to thank the Rutgers Research Council for funding support and the SSEL at The California Institute of Technology for lab support.

References

Andreoni, J., & Miller, J. (2000). Giving according to GARP: an experimental test of the rationality of altruism. Working Paper.

Bolton, G., Brandts, J., & Ockenfels, A. (2005). Fair procedures: evidence from games involving lotteries. *The Economic Journal*, *115*, 1054–1076.

Bolton, G., & Ockenfels, A. (1998). Strategy and equity: an ERC-analysis of the Güth–van Damme game. *Journal of Mathematical Psychology*, *42*, 215–226.

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

Camerer, C. F. (2003). *Behavioral game theory: experiments in strategic interaction*. Princeton: Princeton University Press.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, *117*, 817–869.

Dana, J., Weber, R., & Kuang, J. X. (2004). Exploiting 'Moral Wriggle Room': behavior inconsistent with a preference for fair outcomes. Working Paper, Carnegie Mellon University.

Elster, J. (1998). Emotions and economic theory. *Journal of Economic Literature*, *36*, 47–74.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review*, *94*(4), 857–869.

Fisman, R., Kariv, S., & Markovits, D. (2005a). Individual preferences for giving. Working Paper, UC Berkeley.

Fisman, R., Kariv, S., & Markovits, D. (2005b). Pareto damaging behaviors. Working Paper, UC Berkeley.

Fisman, R., Kariv, S., & Markovits, D. (2005c). Distinguishing social preferences from preferences for altruism. Working Paper, UC Berkeley.

Forsythe, R., Horowitz, J., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*, 347–369.

Güth, W., & van Damme, E. (1998). Information, strategic behavior and fairness in ultimatum bargaining: an experimental study. *Journal of Mathematical Psychology*, *42*, 227–247.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. L. (1994). Preferences, property rights and anonymity in bargaining games. *Games and Economic Behavior*, *7*, 346–380.

Hume, D. (1740). *Treatise on human nature*. London: J.M. Dent, 1939.

Kagel, J., & Wolfe, K. (2001). Testing between alternative models of fairness: a new three person ultimatum game. *Experimental Economics*, *4*, 203–220.

Karni, E., & Safra, Z. (2002a). Individual sense of justice: a utility representation. *Econometrica*, *70*, 263–284.

Karni, E., & Safra, Z. (2002b). Intensity of the sense of fairness: measurement and behavioral characterization. *Journal of Economic Theory*, *105*, 318–337.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *The American Economic Review; Papers and Proceedings*, 426–432.

Rawls, J. (1963). The sense of justice. *Philosophical Review*, *72*, 281–305.

Rawls, J. (1971). *A theory of justice*. Cambridge: Harvard University Press.

Romer, P. M. (2000). Thinking and feeling. *The American Economic Review; Papers and Proceedings*, 439–443.

Smith, A. (1759). The theory of moral sentiments (new ed.), D. D. Raphael & A. L. Macfie (Eds.). Oxford: Oxford University Press, 1976.

Smith, V. L. (1976). Experimental economics: induced value theory. *American Economic Review*, *66*(2), 274–279.

Sopher, B., & Narramore, M. (2000). Stochastic choice in decision making under risk: an experimental study. *Theory and Decision*, *48*, 323–350.