Friedman, Harsanyi, Rawls, Boulding - or Somebody Else?

An Experimental Investigation of Distributive Justice

by Stefan Traub, Christian Seidl, Ulrich Schmidt and Maria Vittoria Levati

C|A|U

Christian-Albrechts-Universität Kiel

Department of Economics

Economics Working Paper

No 2003-03

FRIEDMAN, HARSANYI, RAWLS, BOULDING – OR SOMEBODY ELSE?

An Experimental Investigation of Distributive Justice

Stefan Traub[a*], **Christian Seidl**[a,], **Ulrich Schmidt**[b], **Maria Vittoria Levati**[c]

[a] *Institut für Volkswirtschaftslehre, Abteilung für Finanzwissenschaft und Sozialpolitik, Universität Kiel, D–24098 Kiel, Germany*
[b] *Institut für Volkswirtschaftslehre, Abteilung für Finanzmarkttheorie, Universität Hannover, D–30167 Hannover, Germany*
[c] *Max–Planck–Institut zur Erforschung von Wirtschaftssystemen, Abteilung Strategische Interaktion, D–07745 Jena, Germany*

revised version, May 2003

Abstract. This paper investigates distributive justice using a fourfold experimental design: The ignorance and the risk scenarios are combined with the self–concern and the umpire modes. We study behavioral switches between self–concern and umpire mode and investigate the goodness of ten standards of behavior. In the ignorance scenario, subjects became on average less inequality averse as umpires. A within–subjects analysis shows that about one half became less inequality averse, one quarter became more inequality averse and one quarter left its behavior unchanged as umpires. In the risk scenario, subjects become on average more inequality averse in their umpire roles. A within–subjects analysis shows that half of them became more inequality averse, one quarter became less inequality averse, and one quarter left its behavior unchanged as umpires. As to the standards of behavior, several prominent ones (leximin, leximax, Gini, Cobb-Douglas) experienced but poor support, while expected utility, Boulding's hypothesis, the entropy social welfare function, and randomization preference enjoyed impressive acceptance. For the risk scenario, the tax standard of behavior joins the favorite standards of behavior.

JEL classification: D31, C91, D63
Keywords: Distributive justice, income distributions, veil of ignorance

*Corresponding author: Institut für Volkwirtschaftslehre, Universität Kiel, Olshausenstr. 40, D–24098 Kiel, Germany; Tel.: + 49–431–880–3188; fax: + 49–431–880-4621. *E–mail address: traub@bwl.uni–kiel.de*

1 Introduction

When judged from under a veil of ignorance, income distributions have a striking similarity with lotteries (see, for example, Dahlby, 1987). However, for the evaluation of income distributions, it is often argued that individuals develop social responsibility and would thus, other than with lotteries, exhibit both a risk component and an altruism component of their behavior (Cowell and Schokkaert, 2001, p. 947). This argument has something in it when income distributions are juxtaposed to lotteries the payoffs of which represent only extra incomes. But when lottery payoffs form the bulk of individuals' financial assets, social responsibility of the lucky ones may well befall their behavior and introduce altruism. Thus, to associate selfish behavior with lotteries and social responsibility with income distributions is more a matter of different framing rather than different attitudes.

This paper investigates how the perception of distributive justice of income distributions shifts for two different roles of evaluators when they face two different information scenarios. Varying somewhat Knight (1921), we distinguish between *ignorance* and *risk*. Under risk information, agents know both the possible incomes and their probability distributions. Ignorance assumes that only the set of possible incomes is known, while any probability information is unavailable. Evaluators may judge the distributive justice of income distributions under two roles. First, the evaluator is asked to imagine that that (s)he becomes an income recipient within his or her most favored income distribution after the veil of ignorance has been lifted. We call this environment the *self–concern mode*. Second, the evaluator is assured that (s)he remains an outside observer after the veil of ignorance has been lifted without any stakes in the game. We call this environment the *umpire mode*.

Empirical research on distributive justice can focus on various aspects.

For instance, the pioneering work by Yaari and Bar–Hillel (1984) studies the just division of a commodity basket for different background contexts. Later empirical research on distributional justice extended the frame of reference to *accountability, efficiency,* and *needs* (for a joint study of these aspects see Frohlich and Oppenheimer, 1994; Konow, 2001). The accountability principle requires that subjects' entitlements should vary only with respect to discretionary variables, for which the subjects can be held responsible (see Roemer, 1996, chapters 5–8; Kolm, 1998; Fleurbaey, 1998). In particular, subjects' entitlements should be responsive of effort (see Schokkaert and Overlaet, 1989; Schokkaert and Capeau, 1991; Schokkaert and Devooght 1995). The efficiency principle is concerned with the absolute size of the cake, in other words, with the incentives to maximize the object which is to be distributed. It is much the focus of traditional micro–economics. The needs principle requires that a just distribution should meet each individual's basic requirements for life. It carries over to problems of poverty (see Cowell 2003).

Most empirical research on distributive justice has been done in terms of surveys under the umpire mode. Respondents were asked for their judgement on stories depicting the respective situations. They were like spectators, not actors, of a plot on a fictitious stage.

The present research avails itself of an experiment with material incentives. Subjects' roles are systematically varied between the self–concern mode and the umpire mode. Both modes are probed under the ignorance scenario and the risk scenario, respectively. We investigate systematic shifts in the perception of justice for ten standards of behavior which are considered common expressions of distributional justice. Due to the character of our research as an experiment, we had to restrict our design to purely distributional problems. The inclusion of effort into our experiment would have far transgressed the confines and financial possibilities of our study.

Our subjects had to work in isolation. This means that no discussions were allowed, which might have worked in the direction of agreement on particular standards of behavior (see Miller 1992). This would have shifted the focus of our experiment to the analysis of deliberative democracy. Though this commands interest of its own, an experiment has to restrict attention to a limited number of variations. Monitoring the dynamics of convergence of subjects' perceptions of distributive justice requires a different experimental design.

Section 2 of this paper presents the experimental design, Section 3 discusses thew major possible standards of behavior. Section 4 contains the results of our experiment, and Section 5 concludes.

2 The Experimental Design

2.1 Outline of the Experiment

Combining the self–concern and the umpire modes on the one hand with the ignorance and risk scenarios on the other, gives us the following fourfold pattern of our experimental design:

i) Self–concern mode under ignorance.

ii) Umpire mode under ignorance.

iii) Self–concern mode under risk.

iv) Umpire mode under risk.

The experiment used two sets of stimuli, income sets for the ignorance scenario, income distributions for the risk scenario. Subjects dealt at first with the ignorance scenario, then with the risk scenario. Each scenario consisted of a self–concern part and an umpire part. The instructions given to the subjects are to be found in the Appendix.

61 subjects participated in the experiment, all of them students[1] of the University of Kiel, mostly students of economics, but also students of the business and law schools. In subjects' responses, we did not find gender biases. The experiment was administered in one session which lasted for two hours. All subjects participated in all four parts of the experiment. First, they received an instruction on the experimental design and the payoff procedure the functioning of which is explained below. Then the subjects had to complete four forms by stating their complete preference orderings of the income sets and the income distributions used as stimuli. All four forms were collected and payoffs were effectuated. We chose this procedure to avoid income and information effects of payments.

Payments were worked off for each part separately. A form was drawn at random, the respective payoff procedure was applied, and the subject was immediately paid in cash. The forms of subjects who had received a payoff were not replaced for the particular part of the experiment. Subjects (save the umpires) were, however, not excluded from gaining payments in other parts of the experiment. Due to our budget constraint, we continued payoffs for each of the four parts of the experiment until a ceiling of 500 Deutschmarks of aggregate payoffs was exceeded. This was accomplished

[1]Schokkaert and Capeau (1991) compared the results of a survey with student subjects with the results of a subset of this survey collected from a representative sample of the Flemish working population. They observed rather robust results.

by keeping records of the payments effectuated for the respective part of the experiment. We discontinued payoffs after their sum had exceeded 500 Deutschmarks. Note that no payoffs were pruned: Our aggregate payoff for the whole experiment was 2112 Deutschmarks, which amounts to an average of 35 Deutschmarks per subject. The minimum payoff was zero, the maximum payoff actually made to a subject was 250 Deutschmarks.

2.2 The Stimuli

Stimuli were nine income sets for the ignorance scenario and twelve income distributions for the risk scenario. Subjects received two envelopes with nine and twelve slips of cardboard, respectively, each coded by a symbol to avoid ordering effects triggered by the experimental design. Tables 1 and 2 display a synopsis of the stimuli. The ordering and the numbers in Tables 1 and 2 are only introduced to facilitate reference to the respective income set or income distribution in this paper.

Insert Tables 1 and 2 about here

The nine stimuli of the ignorance scenario show on each slip sets of eligible entries in income distributions (corresponding to reasonable annual incomes in Deutschmarks). Subjects were told that the eventual income distributions were made up only by using components of these sets and no other components must crop up. Moreover, they were told that not all of the components of these sets need to enter the ensuing income distribution. The twelve stimuli of the risk scenario show on each slip an income distribution consisting of exactly five entries representing income quintiles.

Concerning the ignorance scenario, which was mainly adopted to study maximin justice, we were keen to destroy any probability connotation because Rawls' difference principle looses its edge whenever probability information is available and subjects are not absolutely risk averse. This made it necessary to stress the difference between income sets and income distributions in the very notation of the stimuli. For instance, the income distribution resulting from the set 2 in Table 1 is equivalent to the income distribution 1 in Table 2. Had we used the stimuli of the risk scenario also for the ignorance scenario, it could have been misleading to tell subjects that it is one time an income set and the other time an income distribution. Using the principle of insufficient reason, many subjects might have inferred that the five entries in the income sets would occur in exactly one fifth of all cases, thus destroying our efforts to generate a Rawlsian environment. Even this precaution seems not to have hindered some subjects to associate artificially made up probabilities with the entries in the income sets.

Subjects were required to state complete preference orderings of the nine income sets and the twelve income distributions under the self–concern mode and the umpire mode.[2]

[2]Stating complete preference orderings requires much attention on the part of subjects. Asking the subjects for pairwise comparisons, however, would have meant to ask subjects for 102 pairwise preference comparisons instead of but two complete preference orderings. Subjects were simply asked to rank the income sets or income distributions according to their preferences. We did not specify whether they should express their preferences as personal or social preferences because we intended to learn their preferences as a function of their particular *roles* assumed under the self-concern and umpire mode, respectively.

2.3 The Payoff Procedure

The entries in the stimuli were transformed into payments by dividing the amounts in the stimuli by 2000 (roughly equal to the amount of working hours per year).

For payments made under the ignorance scenario, subjects were told that one payoff per slip had been chosen by a neutral person, to wit, our secretary, in advance in an arbitrary way. The neutral person's choices were recorded on a transparency and the transparency was deposited in a sealed envelope. Recall, that, for mimicking Rawls' difference principle, we were keen to destroy any probability connotation. Indeed, our secretary was in no way knowledgeable about our experiment. We simply asked her to arbitrarily touch one number in any set.

Psychological research has shown that subjects seem to follow some causal regularity patterns when they judge or construct probabilities. After ten odd numbers produced by a fair roulette wheel, they tend to estimate the probability of an even number at the next spin as higher than one half. In a commercial lottery with five winning numbers, the winning set $\{1, 2, 3, 4, 5\}$ is considered as less probable by most subjects than five numbers taken from a table of random numbers. Therefore, we assumed that a deliberately arbitrary choice of numbers might best eliminate probability connotations. Had *we* made this choice, subjects would have assumed a tendency towards small payoffs. So we chose our secretary and told our subjects truthfully that she had chosen the numbers without knowing for what purpose she did that.[3] The neutral person's choices were recorded on a transparency and

[3] She had made the following choices (first we indicate the income set, then the chosen entry): 1: 110,000; 2: 60,000; 3: 60,000; 4: 150,000; 5: 30,000; 6: 20,000; 7: 60,000; 8: 250,000; 9: 40,000.

the transparency was deposited in a sealed envelope. Before effectuating payment, the transparency was taken off the envelope and was put on an overhead projector.

For the *self–concern mode* of the ignorance scenario, a subject was drawn at random. Then the subject drew a ball from an urn containing 9 balls bearing a 1 (first rank), 8 balls bearing a 2 (second rank), etc., and, finally, 1 ball bearing a 9 (ninth rank). This procedure was adopted to induce subjects to carefully determine their ordering of income sets. The more preferred income sets had thus a higher chance to be chosen for payoff. The number on the ball drawn determined the subject's respectively ranked slip, which was then selected and payoff was effectuated in cash according to the pre–determined entries as shown on the transparency. Subjects were thoroughly informed about this procedure.

The *umpire mode* of the ignorance scenario followed the same procedure with one important change: *Subjects were instructed to act as impartial umpires who were asked for their advice without any personal involvement.* To this end, the experimenters drew an umpire at random before any payoff was made. The umpire was called to the fore and presented to the audience. The umpire's ranking then determined the payoffs of all other subjects in this part of the experiment. A subject was again drawn at random, drew one ball from the same urn as above, but now the umpire's ranking of income sets was applied instead of the respective subject's ranking. The umpire himself or herself, however, was excluded from any chance to get a payoff. Subjects were again thoroughly informed about this procedure. Thus, in the umpire mode, subjects were aware that, when their ordering of income sets would ever become effective, they themselves would forgo any payoff in this part of the experiment. However, when they were not drawn to become the umpire and could thus participate in the payoffs, their ordering of income sets would become meaningless for the determination of payoffs. This procedure required subjects to feel really as umpires for this part of the experiment. They compiled their orderings of income sets without any own stakes in their outcomes. We chose this experimental design to induce subjects to behave as impartial social planners.

Payment for the risk scenario continued immediately after the payments of the two modes of the ignorance scenario were effectuated.

In the *self–concern mode*, payoffs followed the above procedure (with a correspondingly adapted urn) save that subjects' payoffs were not predetermined, but resulted from subjects' assignments to income quintiles according to a uniform distribution. In fact, we used a wheel of fortune with five equally likely fields in order to assign the subjects to one of the income quintiles. For the *umpire mode*, subjects' own rankings were replaced by the umpire's rankings, the umpire being again excluded from any payoff.

3 Standards of Behavior

3.1 Epistemology of Standards of Behavior

Both the self-concern and the umpire modes rely on distinguishable epistemological roots in philosophy and economics.

The self–concern approach assumes that individuals reveal their preferences for income distributions from under a veil of ignorance, and, after the veil of ignorance has been lifted, become members of their most favored society. This approach was pioneered by Friedman (1953). He argued that income distributions result from *deliberate choices* of agents facing sundry income risks from their decisions regarding jobs and investment. Friedman used expected utility theory assuming full knowledge of the probability distribution of possible incomes on the part of the evaluator. In this he was followed by Strotz (1958, 1961). Kanbur (1979, 1982) extended the Friedman model to a general–equilibrium framework including taxation.

Harsanyi (1953, 1955) developed similar ideas. He, too, relied on expected utility, endowed, however, with a uniform distribution of all eligible incomes.[4] Related ideas were expressed by Vickrey (1945, 1960, 1961), Fleming (1952), Goodman and Markovitz (1952), Dworkin (1981), Dahlby (1987), Kolm (1985, 1998), Fleurbaey (1998), and many others. Diamond (1967) objected that a Harsanyi–type social welfare function cannot exhibit randomization preference, which may be considered a requirement of ex–ante fairness of income distributions. Epstein and Segal (1992) have shown that quadratic social welfare functions satisfy randomization preference.

Rawls (1958, 1971) proposed maximin justice as an equity norm of distributions within the individual–choice approach. According to his difference principle, inequalities in a society are justified as long as they improve the lot of the worst–off agent in the society.[5] For Rawls the bliss of equity is achieved when the lot of the worst–off person cannot be improved any more. Maximin justice is sustained either by the assumption of complete probability ignorance or by the assumption of extreme risk aversion.

[4]Note that Harsanyi's theory is more comprehensive as it encompasses also aspects of group dynamics. These aspects were beyond the scope of our experiment.

[5]We confine our test of the Rawlsian Theory of justice to his difference principle (maximin justice), and reduce his "primary goods" to the income dimension only. Our experiment is simply not rich enough to test the most extensive basic liberty as well as equal access to positions and offices. For an alternative method to test maximin justice see Gaertner, Jungeilges, and Neck (2001). Their results suggest that more prosperous societies are more inclined towards maximin in directing scarce funds to the education of retarded rather than talented children.

Summarizing, Friedman assumes full knowledge of the probabilities of future income positions, Harsanyi, employing the principle of insufficient reason, assumes that all possible incomes are equiprobable, and Rawls assumes complete uncertainty for everything beyond the set of possible incomes.

The umpire approach assumes that some outside judge or observer, a social planner, or some impersonal social welfare function, evaluates the equity of income distributions. The rub of this approach is the *lack of any personal involvement*, that is, the judge, the outside observer, the social planner, or the originator of the social welfare function does not become a member of the society the equity of whose income distribution he or she is going to evaluate.

This approach was pioneered by Dalton (1920) and Atkinson (1970). Important contributions were made by Blackorby and Donaldson (1978), Cowell (1985), and Cowell and Kuga (1981). Major textbooks in this direction are Chakravarty (1990), Cowell (1995), and Lambert (1993). Boulding's (1962, p. 83) proposition that "society lays a modest table at which all can sup and a high table at which the deserving can feast" belongs originally to this approach, although it may as well be put in a framework of self–concern.

It may be immediately gathered from the description of this experiment that its ignorance scenario was devised to mimic a truly Rawlsian setting, while its risk scenario was devised to mimic a truly Friedman–Harsanyi setting. In the ignorance scenario, we tried to extinguish all probability connotations, while the probability part of income distributions was particularly emphasized in the risk scenario. As several scholars (see, for example, Alves and Rossi, 1978; Curtis, 1979; Frohlich et al., 1987a, 1987b; Frohlich and Oppenheimer, 1990; Mitchell et al., 1993) observed but little support for Rawls's difference principle as well as for the Friedman–Harsanyi model of expected utility, but strong support for Boulding's hybrid principle (maximizing average income or expected utility while observing a floor constraint) instead, we were curious to look for the performance of Boulding's principle, which harbors aspects of both scenarios of this experiment. Beyond that, the main focus of our research was directed to investigate differences in subjects' behavior under the self–concern mode and under the umpire mode.

In particular, we were interested to see whether subjects exhibit social responsibility also in the self–concern mode. The explicit reference to the problem framing as one of distributive justice should make subjects aware of the interpersonal concern of their decisions. Even if there is only one more person to individual's self, interpersonal concern has been shown to cause subjects to substantially deviate from individual choices made without the presence of another party. This has been amply demonstrated by Loewenstein, et al. (1989, pp. 437–438). For instance, for the domain of losses, risk seeking in the individual context is largely replaced by risk aversion as well as the desire for equal split in a two–person situation. The same applies for the domain of gains in situations of positive relationship. Interpersonal concern of outcomes was also observed by Curtis (1979, p. 172).

When analyzing our data, we found that subjects were heterogeneous and followed divers heuristics to evaluate income distributions. Moreover, subjects did not put up with a situation of complete ignorance. Instead they devised probability vehicles or ad–hoc heuristics to replace lacking probability information. Some of the heuristics which provide a good explanation of subjects' behavior might look outlandish at first sight. Yet it seems that intuition often outperforms more rational behavior.[6] As we did not confine our

[6]Using computer simulations, Thorngate (1980) has shown that rather crude heuristics, such as the equiprobable heuristic, the minimax regret heuristic, the better than average heuristic, and the probable heuristic (only the last one using probability information at all), perform surprisingly well for the choice of lotteries in selecting the alternative with

test of standards of behavior to a clear–cut set of candidates, we discovered that several standards of behavior showed an impressive performance. For some subjects, however, we did not find plausible behavioral heuristics.

3.2 Taxonomy of Standards of Behavior

This led us to look systematically for standards of behavior which we tested for compliance with our data. The standards of behavior are listed in Table 3. Subjects are supposed to select this one income distribution which maximizes the value of the standard of behavior which is applied by the respective subject. In Table 3, A_{ik}, $k \in K_i$, denotes the entry k in income set or income distribution i, where $i = 1, 2, \ldots, I$, and $I = 9$ for income sets and $I = 12$ for income distributions. Recall that subjects' payoffs were $\frac{A_{ik}}{2000}$. Whenever ordering matters (for the Gini and Tax standards of behavior), all A_i's are arranged in a decreasing order. Zero entries were replaced by 2000 in all logarithmic calculations.

Insert Table 3 about here

When standards of behavior required the employment of utility functions (for EU, PSW, RAP, TAX, and B), we used convex and concave power functions both for infinite bliss and finite agony ($z = 1$), and for finite bliss and infinite agony ($z = -1$), as well as the logarithmic function. As Atkinson's social welfare function is but the α–th root of expected utility, it does not require separate attention, for it yields the same ordering as EU. Note, furthermore, that each entry in an income distribution has a frequency of

the maximum expected value.

one–fifth, so that a subject faces a probability of one fifth to be assigned to each income quintile. Notice that frequencies and assignment probabilities do not make much sense for the income sets of the ignorance scenario. Yet many subjects behaved exactly as if the elements of the income sets in the ignorance scenario had an equal frequency. Some even seemed to have ventured the idea that the values of income sets are best captured by the sum of the utilities of their elements (PSW).

The entries in Table 3 belong to three groups: The first seven entries are *closed–form standards of behavior*. They depict the value of an income set or income distribution as a smooth function of its components; thus, marginal rates of substitution between the components of income distributions exist. Entries eight and nine represent *lexicographic standards of behavior*, and entry ten is a *hybrid standard of behavior*.

Closed–form standards of behavior comprise welfare functions, such as expected utility and Cobb–Douglas welfare functions, welfare functions derived from inequality measures (Blackorby and Donaldson, 1978), quadratic welfare functions which express randomization preference (Epstein and Segal, 1992, p. 700), and evaluation functions which are based on configural weights theory such as the tax model[7] (Birnbaum and Stegner, 1979; Birnbaum and Chavez, 1977, pp. 177–8; Birnbaum, 1999, pp. 41–2). As pseudwelfarism is equivalent to expected utility for homogenous utility functions (such as the power function) and equal frequency, it was tested only for the ignorance

[7] We use a particularly simple form; cf. Birnbaum and Chavez (1997, p. 178). Birnbaum (1999, p. 41) has motivated the name of this model as follows: "This model will be termed the *tax* model to indicate that the weight transferred is a proportion of the weight to be reduced." This means "that weight is transferred among stimuli according to the ranks of the utilities of the outcomes in proportion to the weight of the stimulus that is losing weight."

scenario.

Lexicographic standards of behavior are positional–dictator rules. They confer dictatorship power to individuals in particular positions, for instance, to the worst–off or the best–off individual in a society. Other than closed–form and hybrid standards of behavior, which presuppose unit comparability of utility, lexicographic standards of behavior require but ordinal comparability of utility. These standards comprise Rawls's difference principle (maximin)[8] which was, following a suggestion of Sen (1970, p. 138, note 12), generalized to leximin to avoid ties, and leximax, its counterpart for the best–off individual.

Hybrid standards of behavior are composed of both closed–form and lexicographic standards of behavior. Boulding's principle of maximizing expected utility while observing a floor constraint is just a combination of expected utility and leximin. In a preliminary screening of eligible behavioral patterns, we observed promising hybrid standards of behavior as weighted components of a lexicographic and a closed–form standard of behavior. Yet more comprehensive analyses showed later that all standards of behavior consisting of weighted components of a lexicographic and a closed form standard of behavior are weakly dominated by one of the pure standards of behavior. Therefore, we discarded them from the list of tested standards of behavior.

[8]Note that the maxEmin standard of behavior suggested by Kofler and Zweifel (1988) degenerates for our data to leximin for the ignorance scenario (because of well–known probabilities, this standard of behavior is no candidate for the risk scenario). As we have no data on subjects' a priori probability beliefs, we could only solve the LPI (linear partial information) part of the maxEmin model for any income set i by $\min_{p_k} \sum_{k \in K_i} p_k A_{ik}$, $\sum_{k \in K_i} p_k = 1$, $0 \leq p_k \leq 1$. However, this assigns probability 1 to the minimum income, hence $\min_{p_k} \sum_{k \in K_i} p_k A_{ik} = \min_{k \in K_i} A_{ik}$. Taking then the maximum yields maximin, or, more generally, leximin.

In a separate paper, we made use of the data from this experiment to test Pareto–dominance, Lorenz–dominance, the transfer principle, and decomposability under the two experimental treatments. We found a distinct increase in violations of Pareto–dominance and generalized Lorenz–dominance in the umpire mode, which reflects greater inequality aversion in this mode (cf. Traub et al. 2002). The focus of research in the present paper is, however, directed at a test of standards of behavior.

4 Results

The discussion of the results is arranged in three subsections. We begin with analyzing the ignorance scenario. Then we turn to the data gained from the risk scenario. In these subsections we focus on behavioral shifts between the self–concern mode and the umpire mode in terms of the subjects' compliance with the behavioral standards discussed in Section 3, and in terms of their risk attitudes. Eventually, we compare the subjects' behavior in both scenarios.

4.1 The Ignorance Scenario

Table 4 provides an overview of the rankings of the 9 income sets in the ignorance scenario under the self–concern mode and the umpire mode. Since a Kolmogorov–Smirnov test on the normality of the distribution of ranks rejected the null hypothesis for almost all income sets, we do not only report mean but also median ranks. Under the self–concern mode, income sets (6) and (7) enjoyed most support by the subjects, while income set (8)—the only income set exhibiting a zero entry as one of its components—was bringing up the rear. When moving from the self–concern to the umpire mode, the rank positions of income sets (6) and (7) did not change. However, these income sets had to suffer great losses in both, mean and median rank. On the other hand, there were two big "winners". First, income set (8) gained 5 rank places. Subjects, in their umpire roles, seem to have felt that the possibility of rather high incomes compensates the society for the possibility of zero incomes. However, when possibly affected by a zero income under the self–concern mode, they shy at income set (8). Second, income set (2), which consists of one entry only, gained in terms of mean as well as median rank (though not in terms of rank places). This result revealed an ambivalence of umpires, as they appreciated also the safe side for the society if no big fortunes could have been won. In turn, income set (3), which is strictly dominated by income set (2), lost popularity. These observations were confirmed by two–tailed paired–sample t tests and Wilcoxon rank–sum tests, respectively.

Insert Tables 4 about here

In order to assess the empirical performance of the standards of behavior discussed in the previous section, we adapted a method which was successfully applied by Radzicki (1976, p. 182). First, we computed the *theoretical* ranking of the nine income sets for each standard of behavior. For example, the theoretical ranking implied by the leximin standard of behavior is (2,1,3,5,4,7,6,9,8). For the parametric standards of behavior (EU, ENT, RAP, TAX, B) we recorded all possible theoretical rankings within the feasible parameter set.[9] Note that different standards of behavior may lead to

[9] After computing the theoretical ranking for the lowest feasible parameter value, the parameter was increased in small steps until a new theoretical ranking occurred. If there was more than one change of ranks, the procedure was repeated using a smaller grid size. Otherwise, the new theoretical ranking was recorded, and we looked for the next

the same theoretical ranking of the income sets. For instance, the leximin ranking given above is also obtained for Boulding's standard of behavior with parameters $T = 30000$ and $\alpha = .1$. Altogether, this procedure resulted in 50 different theoretical rankings for the 10 standards of behavior tested in the ignorance scenario.

Second, we computed for every subject Spearman's rank correlation between his or her empirical rank ordering of the nine income sets and any theoretical ranking. This gave us 61 rank correlation coefficients for each theoretical ranking. Table 5 lists for each standard of behavior and for both, the self–concern mode and the umpire mode, the mean and the median rank correlation. For parametric standards of behavior, we picked the parametrization which attained either the maximum mean or the maximum median rank correlation. Parameter values are reported in parentheses right after the name of the respective standard of behavior.

Insert Tables 5 about here

The Friedman test rejected the null hypothesis that the 10 empirical distributions of correlations were drawn from the same sample for both, the self–concern and the umpire mode, as well as for both goodness–of–fit criteria, the maximum mean correlation and the maximum median correlation. The test results are given by $\chi^2(9, N = 61) = 58.68$, $p < .01$ (self concern, mean); $\chi^2(9, N = 61) = 61.06$, $p < .01$ (self concern, median); $\chi^2(9, N = 61) = 63.98$, $p < .01$ (umpire, mean); and $\chi^2(9, N = 61) = 52.19$, $p < .01$ (umpire, median). Computing pairwise Wilcoxon rank–sum test on the equality of two distributions shows that, under the self–concern mode, B, EU, ENT,

theoretical ranking until the upper bound of the parameter range was reached.

and CD significantly outperformed all other standards of behavior according to both criteria ($p \leq .05$; we omit the respective figures to save space). Under the umpire mode, ENT, B, and EU formed the leading group, while CD lost support. Interestingly enough, LMIN was bringing up the rear in all four cases.

Table 5 indicates a shift in the subjects' assessment of the income sets when switching from the self–concern mode to the umpire mode: the between–subjects analysis suggests that subjects behaved less inequality averse under the umpire mode. This pattern is reflected by the optimum values of the parameters α and c of the EU, B and ENT standards of behavior. For instance, under the self–concern mode the highest median correlation for EU was obtained at $\alpha = .26$ while α increased up to .34 under the umpire mode, that is, the shape of the utility function became less concave. In order to check the statistical significance of this result, we computed for every subject his or her individually best matching variant of EU under both, the self–concern and the umpire mode, by varying the risk–aversion parameter α. Then we computed a nonparametric Wilcoxon rank–sum test on the equality of both distributions which, indeed, turned out to be significant (the exact two–tailed significance level was $p = .002$; the medians were .10 and .41, and the means were .29 and .41, respectively). Thus, we can conclude that on average subjects behaved less inequality averse under the umpire mode than under the self–concern mode.

A complementary within–subjects analysis showed that 49% of the subjects behaved less inequality averse (exhibited a higher α under the umpire mode), 28% did not change their behavior in terms of inequality aversion (α was the same under either mode), and only a small group of 23% actually behaved more inequality averse (α decreased). The group behaving less inequality averse when acting as umpires is in conformity with Brickman's (1977) observation that subjects are inclined to endorse more inequality when they have less stakes in the outcome.[10] Recall that umpires have no stakes in the outcome. The other group of subjects behaving more inequality averse when acting as umpires corresponds in its behavior to the findings of Loewenstein et al. (1989, pp. 437–438) and Curtis (1979, p. 172) who observed subjects' dislike even for advantageous inequalities. This applies particularly to the umpire role, in which subjects' concern for own payoff is completely eradicated. Though the former group of subjects is much larger and therefore dominates the between–subjects analysis, the different behavior of both groups is also reflected in Table 4. As mentioned above, income sets (2) and (8) gained most support when switching to the umpire mode. In fact, the popularity growth of the former income set can be attributed to those subjects becoming more inequality averse, as income set (2) is the alternative which not only maximizes the minimum income but also admits no inequality at all. In contrast to this, the group of subjects behaving less inequality averse raised the assessment of alternative (8), the only income set with a zero entry and with the largest range of outcomes.

4.2 The Risk Scenario

Table 6 contains the mean and median rankings of the 12 income vectors in the risk scenario. As compared to the ignorance scenario, the picture seems more clear–cut. Here, the high–payoff, high–risk and high–variance income distributions (8), (9), and (12) lost significantly in mean and median rank in favor of the low–payoff, low–risk and low–variance income distributions (1),

[10]This tendency is more pronounced if ignorance and risk are replaced with certainty about one's own position; see Beckman et al. (2002).

(3), (4), and (5) when moving to the umpire mode. The equal income distribution (1) was the highest winner in mean rank. Table 6 therefore indicates that subjects exhibited on average more inequality aversion as umpires than under self concern. This observation supports the findings of Loewenstein et al. (1989) that the umpire role implies more interpersonal concern for other persons' incomes.

Insert Tables 6 about here

Table 7 contains the assessment of the performance of the behavioral standards in terms of mean and median rank correlations. These were computed using the same procedure as for the ignorance scenario. The risk scenario involved 12 income vectors instead of 9 income sets. Furthermore, PSW was not tested in the risk scenario. We obtained 147 different theoretical rankings for the 9 standards of behavior tested in the risk scenario.

Table 7 shows that, in the risk scenario, EU, RAP, ENT, and TAX outperformed the other standards of behavior. Again, the Friedman test rejected the null hypothesis of the underlying distributions of empirical rank correlations being equal for all standards of behavior under both, the self–concern and the umpire mode, and for both goodness–of–fit criteria: $\chi^2(8, N = 61) = 38.44$, $p < .01$ (self concern, mean), $\chi^2(8, N = 61) = 42.14$, $p < .01$ (self concern, median), $\chi^2(8, N = 61) = 37.94$, $p < .01$ (umpire, mean), and $\chi^2(8, N = 61) = 31.60$, $p < .01$ (umpire, median). Pairwise Wilcoxon rank–sum tests confirmed that the four best matching standards of behavior (EU, RAP, ENT, and TAX) performed significantly ($p \leq .05$) better than all other standards of behavior (we omit the respective figures here to save space). The only exception is B which can catch up with the leading group under the umpire mode when applying the maximum mean criterion.

Insert Tables 7 about here

As opposed to the ignorance scenario, the values of the inequality aversion parameter α decreased when switching to the umpire mode, that is, under the umpire mode subjects on average exhibited more inequality aversion than under the self–concern mode. In order to check for significance of this result, we again computed the individually best matching α for every subject under both modes and performed a Wilcoxon rank–sum test. Under the self–concern mode we obtained a median α of .40 (a mean α of .53) as compared to a median α of .27 (a mean α of .43) for the umpire mode (the exact two–tailed significance level was $p = .028$). Hence, we had to reject the null hypothesis of the two distributions of α coming from the same sample and, therefore, can conclude that subjects on average behaved more inequality averse under the umpire mode than under the self–concern mode.

A within–subjects analysis showed that 48% of the subjects actually exhibited a smaller α under the umpire mode (more inequality aversion), while only 24% had a larger α (less inequality aversion). For the remaining 28% of subjects α was left unchanged when switching between the modes.

4.3 Comparison of Both Scenarios

We computed for all standards of behavior, and within these for all admissible model rankings, the rank correlation coefficients for all 61 subjects. This allowed us to single out the standard of behavior associated with the model variant which exhibited the maximum rank correlation coefficient for any subject. Recall that the best matching standard behavior may not be unique since different standards of behavior (associated with their best model variants) may produce the same maximum value of the rank correlation coefficient. The results of this analysis are shown in Table 8.

Insert Table 8 about here

It is not too surprising that CD, G, LMAX and LMIN performed less good than the parametric standards as the the latter ones had at least one degree of freedom. It is striking, however, that LMIN was the individually best standard of behavior for 16 subjects under the umpire mode of the ignorance scenario although it was bringing up the rear according to our beauty contest of behavioral standards reported in Table 5. The apparent inconsistency can easily be explained by the small subgroup of subjects who became more inequality averse under the umpire mode. Since their attitude is not reflected by the majority, the overall performance of LMIN was very poor as depicted in Table 5. In contrast to its importance in inequality measurement the Gini social welfare function emerged with a rather poor performance from the contest of standards of behavior. The case was still worse for the Cobb–Douglas social welfare function.

PSW, too, does not excel, but, given its lack of logic, did surprisingly well. Subjects seem to have neglect that a greater menu of possible incomes does not increase the total income cake available for a society.

A large group of about one third of subjects shied at zero incomes while making at the same time use of the chances conveyed by higher incomes. Therefore, they settled for the Boulding standard of behavior with $T = 0$, both under the self–concern mode and the umpire mode. The Boulding standard of behavior was a bit less favored in the risk scenario. Here the probabilities of the respective worst outcomes were known and their risk was, therefore, better calculable. Accordingly, some subjects in favor of B in the ignorance scenario switched to EU and RAP in the risk scenario, thus relinquishing their utterly cautious attitude.

Concerning the favorites, namely EU, ENT, RAP, B, and TAX (the last one being a favorite for the risk scenario only), we found that all utility functions were concave (with the exception of TAX for the ignorance scenario and B for the risk scenario) and were of the type "finite agony, infinite bliss". Logarithmic utility functions did not show up. Moreover, most of the best matching utility functions were even more concave than the square–root utility function. This means that subjects were not too sensitive to small and high incomes when determining their standards of behavior (see, in a somewhat different context, Cowell and Flachaire, 2001, Section 3). Higher incomes were given higher weight only if the requirement of a floor constraint was met. This explains why we found $\alpha = 1$ for B in the risk scenario (see Table 7), that is, expected value maximization subject to a floor constraint. Subjects who opted for TAX showed sympathy for a less concave or even a convex (for three cases in the ignorance scenario) utility function, obviously to compensate for the rank–dependent weighting scheme which disfavors high incomes and favors low incomes.

ENT[11] performed about as good as EU did. Note that we observed only the generalized–entropy social welfare function as best matching model variety. The Theil social welfare function ($c = 1$), or the mean–logarithmic–

[11]ENT (Shorrocks, 1980) has the appealing property of decomposability, which is the analogy of the independence condition in risk analysis. See Amiel and Cowell (1999, pp. 15–17).

deviation social welfare function ($c = 0$) never showed up. Moreover, we observed $0 < c \leq 0.55$ for all best model varieties, which means that subjects again were not too sensitive to small and high incomes when determining their standard of behavior (compare Cowell and Flachaire, 2001, Section 3, for a related analysis). RAP enjoyed great support for the risk scenario and somewhat less support for the ignorance scenario. Indeed, in Table 8, RAP was roughly on a par with EU for the ignorance scenario, while Table 7 conveys the impression that it was much inferior. Again, it seems that RAP disposed of a partisan group, and met distinctly less sympathy outside this group. Yet for the risk scenario with its elimination of ignorance about the realization of the entries in the income distribution, RAP gained distinctly more support. In Table 7, it is roughly on a par with EU, in Table 8 it shows somewhat less support. On a whole, RAP commanded considerable support, which is all the more amazing, as it had hardly played a major role in inequality measurement so far. Notice, moreover, that the optimum parameter value for the B's are all such that the κ's are positive. This means that increaing the variance of utility decreases the attractiveness of an income distribution while increasing expected utility increases it. Recall that all utility functions associated with the optimum model variant under RAP are concave and of the "finite agony, infinite bliss" type.

Thus, to come back to the title of our paper, we observe that Friedman, Harsanyi, Boulding and two more hypotheses, to wit, Epstein and Segal (RAP) on the one hand, and Shorrocks (ENT) on the other hand, performed well for both scenarios, while Birnbaum (TAX) performed satisfactorily only for the risk scenario. Rawls, Cobb and Douglas, Gini, and LMAX were the big losers.

The fact that we could not differentiate further between the best four or five standards of behavior can be traced back to the fact that, on the one hand, some behavioral standards generated similar or identical rankings and, on the other hand, the payoff function applied in the experiment was relatively flat[12] For example, in the risk scenario a subject wishing to maximize the expected value of his or her payoff would receive an average payoff of 41 Deutschmarks by stating the respective preference ordering accurately while the average payoff from just stating a random ranking of the 12 income distributions would have been not less than 36 Deutschmarks. This shortcoming, though very hard to detect for the subjects,[13] may potentially have produced some noise in the data.

Finally, we compared the subjects' risk attitudes in terms of the inequality aversion parameter α between the scenarios. Under the self–concern mode, a Wilcoxon rank–sum test significantly rejected the null hypothesis that the distributions of α for the ignorance and the risk scenario, respectively, came from the same sample ($p < .01$). In contrast to this, the null hypothesis could not be rejected under the umpire mode ($p = .85$). Hence subjects on average were least inequality averse under the self–concern mode of the risk scenario and most inequality averse under the self–concern mode of the ignorance scenario. Under the umpire mode subjects exhibited an intermediate degree of inequality aversion irrespective of the scenario.

[12] This point was made by a referee. In order to obtain better results, one could, e.g., replace our linear preference weighting scheme (12, 11,..., 1) by a geometric weighting scheme (2048, 1024, 512,..., 4, 2, 1).

[13] Indeed, subjects meticulously observed the applied linear preference weighting scheme. No one of them ever raised the problem of a too flat payoff function.

5 Conclusions

This paper investigated how the perception of distributive justice of income distributions shifts for two different roles of evaluators when they face two different information scenarios, viz. the *ignorance* and the *risk* scenario. Under risk information, agents know both the possible incomes and their probability distributions. Ignorance assumes that only the set of possible incomes is known, while any probability information is unavailable. Concerning the roles of the evaluators, we exposed them one time to the *self–concern mode*, the other time to the *umpire mode*. Under the self–concern mode, the evaluator becomes an income recipient within his or her most favored income distribution after the veil of ignorance has been lifted. Under the umpire mode, the evaluator is assured that (s)he remains an outside observer after the veil of ignorance has been lifted without any stakes in the game.

The present research availed itself of an experiment with material incentives. Subjects' roles were systematically varied between the self–concern mode and the umpire mode. Both modes were probed under the ignorance scenario and the risk scenario, respectively. We investigated systematic shifts in the perception of justice for ten standards of behavior.

In the ignorance scenario, subjects became on average less inequality averse as umpires. A within–subjects analysis showed that about one half became less inequality averse, one quarter became more inequality averse and one quarter stayed put as umpires. In the risk scenario, subjects became on average more inequality averse in their umpire roles. A within–subjects analysis showed that half of them became more inequality averse, one quarter became less inequality averse, and one quarter did not change its behavior as umpire. As to the standards of behavior, several prominent ones (leximin, leximax, Gini, Cobb-Douglas) experienced but poor support, while expected utility, Boulding's hypothesis, the entropy social welfare function, and randomization preference enjoyed impressive acceptance. For the risk scenario, the tax standard of behavior joins the favorite standards of behavior.

Our observations with regard to the subjects' attitudes towards inequality measured in terms of the inequality aversion parameter point to an important difference in subjects' behavior according to the roles they occupy. The lack of personal involvement under the umpire mode seems to induce a moderate degree of inequality aversion irrespective of the kind of probability information given. In sharp contrast to this, when subjects are personally affected by the realization of a particular income distribution behave on average more inequality averse if no probability information is available. Conversely, they become less inequality averse if they dispose of information about the distribution of outcomes. The impressive performance of Boulding's standard of behavior illustrates that people exhibit a propensity to trade off the chances of admitting more inequality against the risk of being among the worst off in the society. Again the subjects' behavior was affected by the kind of probability information that was given to them. In the ignorance scenario subjects were more inequality averse than in the risk scenario, but they accepted a lower floor constraint. Placing these results in a policy context, we would expect people willing to tolerate more income inequality (implying more personal income risk) if not only the potential consequences of a certain policy or program are known, but their distribution as well. At the same time, people would be willing to safeguard themselves against the risk of being the worst off in the society by calling for a higher subsistence level or poverty line.

Acknowledgements

Financial Support of the European Commission under TMR Contract No. ERBFMRXCT98-0248 is gratefully acknowledged. We are indebted to Serge–Christophe Kolm, Alf Erling Risa, Peter Zweifel, two anonymous referees and an editor of *Social Choice and Welfare* for helpful comments. The usual disclaimer applies. The experimental data can be obtained from Stefan Traub: traub@economics.uni–kiel.de.

References

Alves, W.M., Rossi, P.H., 1978. Who Should Get What? Fair Judgments of the Distribution of Earnings, American Journal of Sociology 84, 541–564.

Amiel, Y., Cowell, F.A., 1999. Thinking about Inequality: Personal Judgment and Income Distributions. Cambridge University Press, Cambridge.

Atkinson, A.B., 1970. On the Measurement of Inequality, Journal of Economic Theory 2, 244–263.

Beckman, S.R., Formby, J.P., Smith, W.J., Zheng, B., 2002. Envy, Malice and Pareto Efficiency: An Experimental Examination, Social Choice and Welfare 19, 349–367.

Birnbaum, M.H., 1999. The Paradoxes of Allais, Stochastic Dominance, and Decision Weights. In: Shanteau, J., Mellers, B.A., Schum, D.A. (Eds.), Decision Science and Technology: Reflections on the Contributions of Ward Edwards. Kluwer, Boston, 27–52.

Birnbaum, M.H., Chavez, A., 1997. Tests of Theories of Decision Making: Violations of Branch Independence and Distribution Independence, Organizational Behavior and Human Decision Processes 71, 161–194.

Birnbaum, M.H., Stegner, S.E., 1979. Source Credibility in Social Judgment: Bias, Expertise, and the Judge's Point of View, Journal of Personality and Social Psychology 37, 48–74.

Blackorby, C., Donaldson, D., 1978. Measure of the Relative Equality and Their Meaning in Terms of Social Welfare, Journal of Economic Theory

18, 59–80.

Boulding, K.E., 1962. Social Justice in Social Dynamics. In: Brandt, R.B. (Ed.), Social Justice. Prentice Hall, Englewood Cliffs, N.J., 73–92.

Brickman, P., 1977. Preference for Inequality, Sociometry 40, 303–310.

Chakravarty, S.R., 1990. Ethical Social Index Numbers. Springer, Heidelberg.

Cowell, F.A., 1985. 'A Fair Suck of the Sauce Bottle' or What Do You Mean by Inequality?, The Economic Record 61, 567–579.

Cowell, F.A., 1995. Measuring Inequality. 2nd ed. Prentice Hall, London.

Cowell, F.A. (ed.), 2003. The Economics of Poverty and Inequality. Edward Elgar, Cheltenham.

Cowell, F.A., Flachaire, E., 2001. Sensitivity of Inequality Measures to Extreme Values. STICERD, London School of Economics, London.

Cowell, F.A., Kuga, K., 1981. Inequality Measurement: An Axiomatic Approach, European Economic Review 15, 287–305.

Cowell, F.A., Schokkaert, E., 2001. Risk Perceptions and Distributional Judgments, European Economic Review 45, 941–952.

Curtis, R.C., 1979. Effects of Knowledge of Self–Interest and Social Relationship Upon the Use of Equity, Utilitarian, and Rawlsian Principles of Allocation, European Journal of Social Psychology 9, 165–175.

Dahlby, B.G., 1987. Interpreting Inequality Measures in a Harsanyi Framework, Theory and Decision 22, 187–202.

Dalton, H., 1920. The Measurement of the Inequality of Incomes, The Economic Journal 30, 348–361.

Diamond, P.A., 1967. Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment, Journal of Political Economy 75, 765–766.

Dworkin, R., 1981. What is Equality? Part 1: Equality of Welfare, Philosophy and Public Affairs 10, 185-246.

Epstein, L.G., Segal, U., 1992. Quadratic Social Welfare Functions, Journal of Political Economy 100, 691–712.

Fleming, M., 1952. A Cardinal Concept of Welfare, The Quarterly Journal of Economics 66, 366–384.

Fleurbaey M., 1998. Equality Among Responsible Individuals. In: Laslier, J.–F., Fleurbaey, M., Gravel, N., Trannoy, A. (Eds.), Freedom in Economics: New Perspectives in Normative Analysis. Routledge, London, 206–234.

Friedman, M., 1953. Choice, Chance, and the Personal Distribution of Income, Journal of Political Economy 61, 277–290.

Frohlich, N., Oppenheimer, J.A., 1990. Choosing Justice in Experimental Democracies with Production, American Political Science Review 84, 461–477.

Frohlich, N., Oppenheimer, J.A., 1994. Preferences for Income Distribution and Distributive Justice: A Window on the Problems of Using Experimental Data in Economics and Ethics, Eastern Economic Journal 20, 147–155.

Frohlich, N., Oppenheimer, J.A., Eavey, C.L., 1987a. Laboratory Results on Rawls's Distributive Justice, British Journal of Political Science 17, 1–21.

Frohlich, N., Oppenheimer, J.A., Eavey, C.L., 1987b. Choices of Principles of Distributive Justice in Experimental Groups, American Journal of Political Science 13, 606–636.

Gaertner, W., Jungeilges, J., Neck, R., 2001. Cross–Cultural Equity Evaluations: A Questionnaire–Experimental Approach", European Economic Review 45, 953–963.

Goodman, L., Markovitz, H., 1952. Social Welfare Functions Based on Individual Rankings, American Journal of Sociology 57, 257–262.

Harsanyi, J.C., 1953. Cardinal Utility in Welfare Economics and in the Theory of Risk–Taking, Journal of Political Economy 61, 434–435.

Harsanyi, J.C., 1955. Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility, Journal of Political Economy 63, 309–321.

Kanbur, S.M., 1979. Of Risk Taking and the Personal Distribution of Income, Journal of Political Economy 87, 769–797.

Kanbur, S.M., 1982. Entrepreneurial Risk Taking, Inequality, and Public Policy: An Application of Inequality Decomposition Analysis of the General Equilibrium Effects of Progressive Taxation, Journal of Political Economy 90, 1–21.

Knight, F.H., 1921. Risk, Uncertainty, and Profit. University of Chicago Press, Chicago.

Kofler, E., Zweifel, P., 1988. Exploiting Linear Partial Information for Optimal Use of Forecasts. With an Application to U.S. Economic Policy, International Journal of Forecasting 4, 15-32.

Kolm, S.–C., 1985. Le contrat social libéral: Philosophie et pratique du libéralisme. Presses Universitaires de France, Paris.

Kolm, S.–C., 1998. Chance and Justice: Social Policies and the Harsanyi–Vickrey–Rawls Problem, European Economic Review 42, 1993–1416.

Konow, J., 2001. Fair and Square: The Four Sides of Distributive Justice, Journal of Economic Behavior and Organization 46, 137–164.

Lambert, P.J., 1993. The Distribution and Redistribution of Income. 2nd ed. Manchester University Press, Manchester.

Loewenstein, G.F., Thompson, L., Bazerman, M.H., 1989. Social Utility and Decision Making in Interpersonal Contexts, Journal of Personality and Social Psychology 57, 426–441.

Miller, D., 1992. Deliberative Democracy and Social Choice, Political Studies 40 (Special Issue), 54–67.

Mitchell, G., Tetlock, P.E., Mellers, B.A., Ordóñez, L.D., 1993. Judgments of Social Justice: Compromises Between Equality and Efficiency, Journal of Personality and Social Psychology 65, 629–639.

Radzicki, J., 1976. Technique of Conjoint Measurement of Subjective Value of Own and Other's Gains, Polish Psychological Bulletin 7, 179–186.

Rawls, J., 1958. Justice as Fairness, Philosophical Review 67, 164–194.

Rawls, J., 1971. A Theory of Justice. Harvard University Press, Cambridge, MA.

Roemer; J.E., 1996. Theories of Distributive Justice. Harvard University Press, Cambridge, MA.

Schokkaert, E., Capeau, B., 1991. Interindividual Differences in Opinions About Distributive Justice, KYKLOS 44, 325–345.

Schokkaert, E., Devooght, K., 1995. The Empirical Acceptance of Compensation Axioms, CES, Public Economics Research Paper No. 45.

Schokkaert, E., Overlaet, B., 1989. Moral Intuitions and Economic Models of Distributive Justice, Social Choice and Welfare 6, 19–31.

Sen, A.K., 1970. Collective Choice and Social Welfare. Holden–Day and Oliver & Boyd, San Francisco.

Shorrocks, A.F., 1980. The Class of Additively Decomposable Inequality Measures, Eonometrica 48, 613–625.

Strotz, R.H., 1958. How Income Ought to be Distributed, A Paradox in Distributive Ethics, Journal of Political Economy 66, 189–205.

Strotz, R.H., 1961. How Income Ought to be Distributed: Paradox Regained, Journal of Political Economy 69, 171–178.

Thorngate, W., 1980. Efficient Decision Heuristics, Behavioral Science 25, 219–225.

Traub, S., Seidl, C., Schmidt, U. 2002. Lorenz, Pareto, Pigou: Who Scores Best? University of Kiel, Economic Working Paper 4/2003.

Vickrey, W., 1945. Measuring Marginal Utility by Reactions to Risk, Econometrica 13, 319–333.

Vickrey, W., 1960. Utility, Strategy and Social Decision Rules, Quarterly Journal of Economics 74, 507–535.

Vickrey, W., 1961. Risk, Utility and Social Policy, Social Research 28, 205–217.

Yaari, M.E., Bar–Hillel, M. 1984. On Dividing Justly, Social Choice and Welfare 1, 1–24.

Appendix: Instructions

At the beginning of the experiment, two closed envelopes containing 9 and 12 slips of cardboard, respectively, were handed over to the subjects. Furthermore, two urns containing 45 and 78 numbered balls, respectively, and a wheel of fortune with five equally likely sectors numbered from 1 to 5 were placed on a table in front of the subjects. The subjects received a sheet of paper with instructions. The instructions were also read out aloud, and the subjects were given some time to study them on their own, and to ask questions.

> *Dear participant!*
>
> *We would like to thank you for participating in our experiment.*
>
> *In this experiment, you will be asked to rank different income distributions according to their desirability. There will be four different setups. When all decisions have been made, we will draw—for each setup separately—participants randomly and pay them off according to their decisions until the sum of payoffs exceeds a budget of 500 Deutschmarks in each setup.*
>
> *For a participant drawn, the payoff is determined as follows: Assume that you ranked N different income distributions. Now, a ball is drawn from an urn which contains N balls with number 1, $N-1$ balls with number 2, $N-2$ balls with number 3, and so on, and 1 ball with number N. Your payoff is then determined by the income distribution for which your ranking and the number of the ball drawn correspond. Further details depend on the setup and will be explained on separate questionnaires.*

Please mark each questionnaire with your name, and mark with a cross whether your are male or female.

After reading out these instructions, the first questionnaire was handed over to the participants and, again, read out aloud to the subjects.

In the first envelope, you will find 9 sets of components (annual net incomes in Deutschmarks) which make up possible income distributions. Please, rank these sets according to your preferences. If you are drawn for receiving a payoff, one of the sets will be drawn following the above procedure. Then you will receive 1/2000 *of one of the components as your payoff. For each set, the respective income has been chosen arbitrarily by our secretary. Enter your preference order of the income sets using the symbols displayed on the slips of cardboard in the table below.*

When the first part was completed, questionnaire 2 was distributed among the subjects.

In the second envelope, you will find 12 income distributions. The income recipients are split into 5 equally sized groups, each amounting to 20% of the population. Please, rank the income distributions (annual net incomes in Deutschmarks) according to your preferences. If you are drawn for receiving a payoff, in the first step, you will be assigned to one of the five income groups with a 20% probability each. In the second step, an income distribution will be drawn randomly and you receive 1/2000 *of the income corresponding to your income group. Enter your preference order of the income distributions using the symbols displayed on the slips of cardboard in table below.*

The third part of the experiment was introduced by the following questionnaire:

> *Consider the 9 sets of components, which make up possible income distributions, from envelope 1 again. Please, rank these sets according to your preferences. Note, however, that one participant is drawn randomly at the end of the experiment and becomes an umpire. The name of the umpire and his or her decisions will be made public. The rank order of the umpire determines the probability of one of the sets being chosen for the whole group of participants. If you are drawn for receiving a payoff, the set chosen according to the umpire' preferences determines your payoff. You will get* 1/2000 *of one of the components as your payoff. For each set, the respective income has been chosen arbitrarily by our secretary. The umpire is excluded from getting any payoff. Enter your preference order of the income sets using the symbols displayed on the slips of cardboard in the table below.*

Eventually, questionnaire 4 was handed over to the subjects.

> *Consider the 12 income distributions from envelope 2 again, where the income recipients are split into 5 equally sized groups of 20% of the population. Please, rank the income distributions according to your preferences. Note, however, that one participant is drawn randomly at the end of the experiment and becomes an umpire. The name of the umpire and his or her decisions will be made public. Now, the rank order of the umpire determines the probability of one of the income distributions being chosen for the whole group of participants. If you are drawn for receiving*

a payoff, in the first step, you will be assigned to one of the five income groups with a 20% probability each. In the second step, an income distribution will be drawn according to the umpire's preferences and you receive 1/2000 *of the income corresponding to your income group. The umpire is excluded from getting any payoff. Enter your preference order of the income distributions using the symbols displayed on the slips of cardboard in the table below.*

Tables

Table 1

Stimuli for ignorance scenario

No.	Symbol	Name	Income set
1	□	square	{59,000 110,000}
2	⋄	diamond	{60,000}
3	◯	circle	{40,000 45,000 50,000 55,000 60,000}
4	+	cross	{30,000 150,000}
5	$\mathbb{X}$	swords	{30,000 180,000}
6	△	triangle	{20,000 50,000 100,000 150,000 220,000}
7	▽	giveaway	{20,000 60,000 100,000 160,000 220,000}
8	—	horline	{0 100,000 220,000 250,000}
9	|	verline	{10,000 20,000 30,000 40,000 45,000 50,000 55,000 60,000 80,000 90,000 100,000}

Table 2

Stimuli for risk scenario

No.	Symbol	Name	Income distribution
1	$\square$	square	(60,000 60,000 60,000 60,000 60,000)
2	$\diamond$	diamond	(50,000 55,000 60,000 65,000 70,000)
3	$\bigcirc$	circle	(40,000 50,000 60,000 70,000 80,000)
4	$+$	cross	(40,000 40,000 60,000 80,000 80,000)
5	$\bowtie$	bowtie	(40,000 60,000 60,000 60,000 80,000)
6	$\mathbb{X}$	swords	(10,000 20,000 60,000 100,000 110,000)
7	$\triangle$	triangle	(10,000 60,000 60,000 60,000 110,000)
8	$\bigtriangledown$	giveaway	(70,000 70,000 100,000 110,000 120,000)
9	—	horline	(70,000 70,000 70,000 90,000 180,000)
10	$\vert$	verline	(15,000 15,000 100,000 110,000 120,000)
11	$\overset{\triangledown}{\triangle}$	sandglas	(15,000 15,000 70,000 90,000 180,000)
12	$\boxtimes$	crossbox	(0 60,000 80,000 250,000 250,000)

Table 3
Tested standards of behavior

Standard of behavior		Formula	Parameter range
EU	Expected Utility[a,b]	$\max_{i \in I} \left\{ \frac{\sum_{k \in K_i} z \left(\frac{A_{ik}}{2000} \right)^{\alpha}}{\# K_i} \right\}$, $z = \begin{cases} -1 & \alpha < 0 \\ 1 & \alpha \geq 0 \end{cases}$	$\alpha \in [-2, 2]$
PSW	Pseudowelfarism[a,c]	$\max_{i \in I} \left\{ \sum_{k \in K_i} z \left(A_{ik} \right)^{\alpha} \right\}$	$\alpha \in [-2, 2]$
CD	Cobb–Douglas	$\max_{i \in I} \left\{ \left[\prod_{k \in K_i} \left(\frac{A_{ik}}{2000} \right)^{\frac{1}{\# K_i}} \right] \right\}$	—
G	Gini	$\max_{i \in I} \left\{ \frac{1}{2000 (\# K_i)^2} \sum_{k=1}^{\# K_i} (2k - 1) A_{ik} \right\}$	—
ENT	Entropy	$\max_{i \in I} \left\{ \frac{\bar{A}_i}{2000} \left\{ 1 - \frac{1}{c^2 - c} \left[\frac{1}{\# K_i} \sum_{k \in K_i} \left(\frac{A_{ik}}{\bar{A}_i} \right)^c - 1 \right] \right\} \right\}$	$c \in [-100, 100]$, $c \neq 0, 1$
		$\max_{i \in I} \left\{ \frac{\bar{A}_i}{2000} \left[1 + \frac{1}{\# K_i} \sum_{k \in K_i} \ln \left(\frac{A_{ik}}{\bar{A}_i} \right) \right] \right\}$	$c = 0$
		$\max_{i \in I} \left\{ \frac{\bar{A}_i}{2000} \left[1 - \frac{1}{\# K_i} \sum_{k \in K_i} \left(\frac{A_{ik}}{\bar{A}_i} \right) \ln \left(\frac{A_{ik}}{\bar{A}_i} \right) \right] \right\}$	$c = 1$
RAP	Randomization Preference[a,d]	$\max_{i \in I} \left\{ \# K_i (\# K_i - 1 + B) \left\{ [\mathrm{E}(u_i)]^2 - \kappa_i \mathrm{var}(u_i) \right\} \right\}$ $u_i = z \left(\frac{A_{ik}}{2000} \right)^{\alpha}$, $\kappa_i = \frac{1 - B}{\# K - 1 + B}$	$\alpha \in [-2, 2]$ $B \in (-4, 1)$
TAX	Tax Model[a]	$\max_{i \in I} \left\{ \sum_{k=1}^{\# K_i} k z \left(\frac{A_{ik}}{2000} \right)^{\alpha} \right\}$	$\alpha \in [-2, 2]$
LMIN	Leximin	$\max_{i \in I} \left\{ \min_{k \in K_i} \{ A_{ik} \} \right\}$, $\min_{k \in K_j} \{ A_{jk} \} > \min_{k \in K_i} \{ A_{ik} \} \; \forall \; j \neq i$	—
LMAX	Leximax	$\max_{i \in I} \left\{ \max_{k \in K_i} \{ A_{ik} \} \right\}$, $\max_{k \in K_j} \{ A_{jk} \} < \max_{k \in K_i} \{ A_{ik} \} \; \forall \; j \neq i$	—
B	Boulding	LMIN if $\max_{i \in I} \left\{ \min_{k \in K_i} \{ A_{ik} \} \right\} \leq T$ EU if $\max_{i \in I} \left\{ \min_{k \in K_i} \{ A_{ik} \} \right\} > T$	$0 \leq T < 250000$

[a]Tested also for logarithmic utility.
[b]Gives same ordering as Atkinson social welfare function.
[c]Tested only for the ignorance scenario.
[d]$B \in (-1, 1)$ for the ignorance scenario.

Table 4
Mean and median rankings of income sets in the ignorance scenario

No.	Income set	Self–concern			Umpire			
		Mean rank	KS–Z[a]	Median rank	Mean rank	KS–Z[a]	Median rank	T test[b] Wilcoxon test[c]
1	{59, 110}	4.11 (3)	*1.30	4	4.25 (3)	*1.31	4	−0.13
		.31			*.28*			.60
2	{60}	5.51 (5)	*1.23	6	4.70 (5)	*1.26	4	*+0.80
		.34			*.38*			**1.85
3	{40, 45, 50, 55, 60}	5.92 (7)	**1.42	7	6.64 (9)	**1.66	7	*−0.72
		.33			*.27*			*1.88
4	{30, 150}	5.93 (8)	**1.46	6	5.49 (7)	1.20	6	+0.44
		.19			*.23*			1.09
5	{30, 180}	5.07 (4)	1.19	5	5.07 (6)	*1.27	5	0
		.23			*.23*			.19
6	{20, 50, 100, 150, 220}	3.48 (2)	**1.65	3	4.15 (2)	**1.58	4	**−0.67
		.23			*.25*			*1.86
7	{20, 60, 100, 160, 220}	2.92 (1)	**1.62	2	3.92 (1)	**1.65	3	**−1.0
		.25			*.33*			**2.32
8	{0, 100, 220, 250}	6.16 (9)	**2.15	7	4.51 (4)	**1.46	4	**+1.65
		.40			*.42*			**3.35
9	{10, 20, 30, 40, 45, 50	5.90 (6)	**1.83	7	6.28 (8)	**1.78	7	−0.38
	55, 60, 80, 90, 100}	*.36*			*.35*			.77

Table note. *$p \leq .10$, **$p \leq .05$. Standard errors in italics.
[a] Z statistic of the Kolmogorov–Smirnov test on normality. Significance levels based on Monte–Carlo simulations.
[b] Mean difference. Two–tailed paired–sample t test.
[c] Z statistic of the Wilcoxon test. Significance levels based on Monte–Carlo simulations.

Table 5
Goodness of standards of behavior—ignorance scenario

Self–concern				Umpire			
Mean correlation		Median correlation		Mean correlation		Median correlation	
B ($T = 0, \alpha = .1$)	.4051	B ($T = 0, \alpha = .2625$)	.5166	ENT ($c = .55$)	.3136	ENT ($c = .4$)	.4166
EU ($\alpha = .1$)	.4051	EU ($\alpha = .2625$)	.5166	B ($T = 0, \alpha = .3875$)	.2956	B ($T = 0, \alpha = .3375$)	.4000
ENT ($c = .1$)	.3945	ENT ($c = .4$)	.5000	EU ($\alpha = .3875$)	.2956	EU ($\alpha = .3375$)	.4000
CD	.3945	CD	.3500	LMAX	.2387	LMAX	.2833
LMAX	.1781	RAP ($B = -.76, \alpha = .6$)	.1666	G	.2387	G	.2833
G	.1781	PSW ($\alpha = .4$)	.1666	CD	.2180	CD	.2666
RAP ($B = -.76, \alpha = .1$)	.1475	LMAX	.1666	PSW ($\alpha = 1.725$)	.1393	PSW ($\alpha = 1.525$)	.2166
PSW ($\alpha = .1$)	.1475	G	.1666	TAX ($\alpha = 1.95$)	.0185	RAP ($B = -.99, \alpha = .7$)	.1166
TAX ($\alpha = .1$)	.1475	TAX ($\alpha = 1.2$)	.1666	RAP ($B = -.99, \alpha = .7$)	.0120	TAX ($\alpha = 1.95$)	.0666
LMIN	.0158	LMIN	.0833	LMIN	-.0136	LMIN	-.0333

Table 6
Mean and Median rankings of income distributions in the risk scenario

No.	Income distribution	Self concern Mean rank	Self concern KS-Z[a]	Self concern Median rank	Umpire Mean rank	Umpire KS–Z[a]	Umpire Median	T test[b] Wilcoxon test[c]
1	(60, 60, 60, 60, 60)	6.95 (5)	**1.45	7	5.62 (3)	*1.22	5	**+1.33
		.43			*.44*			**2.39
2	(50, 55, 60, 65, 70)	6.66 (4)	*1.24	6	6.28 (5)	1.13	6	+0.38
		.38			*.37*			.55
3	(40, 50, 60, 70, 80)	7.61 (9)	1.02	7	6.41 (6)	.84	6	**+1.20
		.30			*.31*			**2.63
4	(40, 40, 60, 80, 80)	7.79 (10)	.84	8	6.77 (8)	1.02	7	**+1.02
		.27			*.29*			**2.73
5	(40, 60, 60, 60, 80)	7.30 (8)	*1.27	7	6.08 (4)	1.06	6	**+1.22
		.28			*.34*			**.264
6	(10, 20, 60, 100, 110)	9.02 (12)	**1.74	10	9.03 (12)	**1.62	10	−0.01
		.38			*.42*			.19
7	(10, 60, 60, 60, 110)	8.46 (11)	1.18	9	7.75 (11)	**1.49	9	+0.71
		.36			*.43*			1.04
8	(70, 70, 100, 110, 120)	2.48 (1)	**2.11	2	4.15 (1)	**1.71	3	**−1.67
		.21			*.45*			**3.28
9	(70, 70, 70, 90, 180)	2.59 (2)	**2.27	2	4.34 (2)	**2.23	2	**−1.75
		.25			*.46*			**3.25
10	(15, 15, 100, 110, 120)	7.07 (6)	**1.41	8	7.46 (9)	1.19	8	−0.39
		.33			*.35*			1.26
11	(15, 15, 70, 90, 180)	7.11 (7)	**1.39	7	7.59 (10)	*1.31	8	−0.48
		.42			*.36*			.84
12	(0, 60, 80, 250, 250)	4.98 (3)	**1.86	3	6.51 (7)	**1.59	5	**−1.53
		.58			*.61*			*1.96

Table note. $^*p \leq .10$, $^{**}p \leq .05$. Standard errors in italics.
[a] Z statistic of the Kolmogorov–Smirnov test on normality. Significance levels based on Monte–Carlo simulations.
[b] Mean difference. Two–tailed paired–sample t test.
[c] Z statistic of the Wilcoxon test. Significance levels based on Monte–Carlo simulations.

Table 7
Goodness of standards of behavior—risk scenario

Self–concern				Umpire			
Mean correlation		Median correlation		Mean correlation		Median correlation	
EU ($\alpha = .4382$)	.5203	EU ($\alpha = .2641$)	.6713	EU ($\alpha = .2563$)	.3715	RAP ($B = -2, \alpha = .625$)	.6153
RAP ($B = .25, \alpha = .5328$)	.5195	RAP ($B = -1.25, \alpha = .4438$)	.6713	RAP ($B = -1.25, \alpha = .425$)	.3715	ENT ($c = .2375$)	.5944
ENT ($c = .4$)	.4939	ENT ($c = .4$)	.6573	ENT ($c = .35$)	.3704	TAX ($\alpha = .4039$)	.5944
TAX ($\alpha = .6785$)	.4939	TAX ($\alpha = .6785$)	.6573	TAX ($\alpha = .5285$)	.3704	EU ($\alpha = .2375$)	.5804
B ($T = 0, \alpha = 1$)	.3743	B ($T = 15000, \alpha = 1$)	.4265	B ($T = 15000, \alpha = 1$)	.3457	B ($T = 15000, \alpha = .625$)	.4737
CD	.3418	CD	.3846	CD	.3291	CD	.4475
LMIN	.3418	LMIN	.3846	LMIN	.3291	LMIN	.4475
G	.3033	G	.2657	G	.0177	G	.0139
LMAX	.3033	LMAX	.2657	LMAX	.0177	LMAX	.0139

Table 8
Individually best matching standards of behavior

	Ignorance		Risk	
	Self concern	Umpire	Self concern	Umpire
B	22	23	14	14
CD	0	0	2	2
ENT	19	21	15	23
EU	14	7	26	22
G	3	7	2	4
LMAX	3	7	2	4
LMIN	8	16	2	2
PSW	9	7	–	–
RAP	14	6	19	20
TAX	2	5	11	11
Sum	94	99	93	102

Note: Includes multiple assignments.