

*Choices of Principles of Distributive Justice in Experimental Groups**

Norman Frohlich, *University of Manitoba*

Joe A. Oppenheimer, *University of Maryland*

Cheryl L. Eavey, *Florida State University*

Experimental methods involving imperfect information are used to generate group choices of principles of distributive justice. Conditions approximating John Rawls's "original position" in *A Theory of Justice* serve as the starting point, and his conjectures are contrasted with those of John Harsanyi. Three "predictions" implicit in the Rawlsian argument are tested: (1) individuals choosing a principle of economic distribution would be able to reach unanimous agreement; (2) they would always choose the same principle; and (3) they would always choose to maximize the welfare of the worst-off individual. Our results indicate that individuals reach consensus, strongly reject the minimax principle, and largely choose what Rawls has called an "intuitionistic" principle. Overwhelmingly, the chosen principle is maximizing the average income with a floor constraint: a principle which is a compromise between those proposed by Rawls and Harsanyi. It takes into account not only the position of the worst-off individual but also the potential expected gain for the rest of society.

One of the most important recent lines of inquiry in political philosophy deals with the question of distributive justice. One branch of this literature has used the game theoretic notion of imperfect information to build upon a tradition of "impartial reasoning" (which itself stems from Kant). These works have subsumed aspects of distributive justice under questions of fair division. John Rawls's *A Theory of Justice*, although the most celebrated of these works, is only one example of such models. (For an alternative formulation, see Harsanyi, 1953, 1955.)

Rawls's conclusions are typical of this genre. He argues that under very specific hypothetical conditions (called an "original position") a set of individuals would *unanimously* choose, as the governing principle of

*We thank the Social Science and Humanities Research Council of Canada, the National Science Foundation (Grant No. SES 82 07807, A01), and our respective universities for their essential support of this research. The additional support of Nuffield College, during our sabbaticals allowed us to sharpen our wits as well as our pencils. The assistance of Joe Aubin, Irvin Boschmann, David Cross, Pam Edwards, Valerie Lehr, Paul Parker, Genie Stowers, and Craig Conners in running the experiments and helping with computer matters was crucial for the completion of the work. Advice from Pat Bond, Ellie Tragakis, Mark Winer, and Gary Miller also helped in the design. Comments by Fred Alford, Dennis McGrath, Rodger Noll, Norman Schofield, Tom Schwartz, Matthew Spitzer, Ric Uslaner, and Oran Young are greatly appreciated by the authors.

distributive justice, to *maximize the welfare of the worst-off* individual in the society. Thus, Rawls's theory (and similar works) concentrate on the *pattern* of distribution and deemphasize questions of entitlements. By contrast, another set of works underscores the role of entitlements in questions of distributive justice. (For a salient example, see Nozick, 1974.)

Both strands of work on justice have generated widespread debate, but the underlying assumptions of the two constructs have not been subject to systematic empirical testing.¹ In this paper we report on a series of experiments explicitly designed as a partial test of theories of justice as fair division, reserving for the future a similar discussion of justice based on entitlements. Our hope is that these efforts will demonstrate the value and feasibility of subjecting aspects of ethical theories to empirical testing.

Inducing Consensus about Distributive Justice via Imperfect Information

If some patterns of distribution are more just than others, how can we identify their ordering? A general procedure has been proposed for doing this: ask individuals to decide upon a principle for dividing goods when they lack knowledge regarding which share they will be allotted. To the extent these individuals are concerned about fairness, their ignorance induces "impartial reasoning." This notion lies at the heart of "justice as fair division" arguments.

Since the most highly developed conditions for inducing preferences for fair division are those of Rawls (1971), we shall briefly outline his argument and show how our experimental design attempts to capture and test components of the generic procedure implicit in his model.

Rawls's Derivation of the Difference Principle

Rational choice is determinate only if constraints are imposed upon the set of alternatives from which individuals are to choose and only if some aspects of individuals' "preferences" are assumed. Thus, the theorist must posit psychological assumptions about the nature of rational choice, the constraints facing the individuals, and the values upon which the choices depend. Rawls does this by means of a "thought experiment" in which he specifies conditions conducive to the "discovery" and selection of a principle of distributive justice. These conditions constitute the "original position" within which the individuals are to choose the terms of contract for the setting up of society. The contract is to include both principles governing the distribution of wealth and income and the set

¹ See such works as Hochschild (1981) and Soltan (1982) for a presentation and review of some preliminary tests of some aspects of this subject.

of individual rights which inhere in Rawls's just society. Although, for Rawls, the moral status of these latter rights takes precedence over the principles of distributive justice, here we consider only the problem of choosing a principle of distributive justice.²

Rawls attempts to utilize characteristics of the original position to *derive* the conclusion that the individuals' deliberations will yield the difference principle as the unanimous choice. In his argument Rawls uses a combination of items: (1) the psychology of humans; (2) the stakes involved in the distributive issue; (3) the restricted level of information available to the individuals; (4) the agenda of principles of distributive justice; and (5) the procedures for the discussion of and voting upon the most preferred principle.

He assumes that the individuals involved are rational and self-interested. In this, his assumption corresponds to traditional microeconomic assumptions and with other models of fair division. Where he differs from others (such as Harsanyi) is in his (somewhat) implicit assumption that individuals are (in the decision context he posits) almost exclusively concerned with the possibility that they may be among the worst off. In the game-theoretic sense, they employ a maximin rule in their choices.³ The stakes posited by Rawls are high: one's life chances and the life chances of all of one's descendants. The only restriction on the range of the stakes is Rawls's proviso that the society in question is to be one of "moderate scarcity."

The restriction on information (the "veil of ignorance") is used to induce impartial reasoning among the individuals. Individuals cannot know their own talents, skills, tastes, nor statuses in the society for which they are making a decision. Nor can the individuals have specific information regarding the nature of that society beyond the knowledge that it is one of moderate scarcity (Rawls, 1971, pp. 18–19, 136–42). Moreover, the individuals are assumed to come to the original position without preconceptions as to what constitutes social welfare, fairness, or "the good."

As to the agenda, Rawls even specifies (p. 124) which principles of distributive justice should be considered. Individuals are charged with discussing (among others) principles which: (1) maximize the welfare of the worst-off individual in the society; (2) maximize the total utility in the

² We do not examine or test the wider issue of how individual rights interact with the choice of distributive principles.

³ Rawls's formulation of the maximin principle of justice is related to the psychological models, or behavioral assumptions, which underlie game theory. Von Neumann and Morgenstern (1944) postulated that when faced with choices under *uncertainty* (i.e., where there is no information regarding the probabilities of relevant outcomes) individuals would choose to maximize their "security level" or "floor." According to Rawls (1971, pp. 154–55) this hypothesis of floor maximizing is the appropriate one to describe behavior in the original position.

society; (3) maximize the average utility in the society (the latter two are identical if societal size is fixed); and (4) a variety of “mixed” principles.

Rawls insists that the decision be preceded by an extensive discussion in which the merits of all of the principles be considered. Furthermore, the decision is to be reached in a “consensus reaching style” with open discussion. Individuals are to continue discussions until they feel comfortable with their decisions. Only after individuals have exhausted their arguments and reached a “reflective equilibrium” is a formal decision to be reached. The rule employed is to be unanimity.

Having set the stage so carefully, Rawls argues that behind the veil of ignorance, individuals’ fears of being placed in the worst-off class would be so great that they would all agree to a distributive rule which maximized the welfare of the worst off. This can be conceptualized as a “maximin rule of distributive justice.” He argues that after careful deliberation the difference principle would be unanimously approved by individuals who had reached a reflective equilibrium.

In summary then, under these idealized conditions, Rawls argues that:

1. Individuals will always reach unanimous agreement.
2. Individuals will always choose the same principle.
3. The principle that will be chosen is the “difference principle”—a distributive principle which maximizes the welfare of the worst-off individual in the society.

Despite its importance in Rawls’s argument and its bearing on related arguments, the original position is an ideal construct: one which may never be realized. The fact that it is an idealization does not, however, mean that the argument is without empirical content. Indeed, one could argue that only insofar as it has an empirical element is (political) ethics more akin to the empirical sciences than it is to an extension of logic and mathematics. Thus, our interest here is *not* in what some have called the *geometry* of morals, but rather in the *physics* of morals.

If Rawls were right (and we dispute that below), individuals in the original position would choose to maximize the welfare of the worst off. As an ethical argument, this would only be *compelling* if—as the ideal were approximated empirically—the behavior of individuals came to approximate what was predicted in the ideal case. That is, the degree to which Rawls is compelling is precisely the degree to which Rawls’s theoretical constructs are robust.⁴

⁴ The conception here is similar to a notion of *robustness* regarding scientific theories. Bridge concepts, such as friction or air resistance, serve to permit measurement of the approximation of ideal constructs such as “vacuum.” As the ideal is approached, the

Was Rawls right? Logically, is the floor the only concern? We believe that individual analyses of choices are richer than a narrow-minded focus on the worst outcome. In previous experiments conducted in different contexts (Miller and Oppenheimer, 1982; Eavey and Miller, 1984; and Frohlich et al., 1984), we observed individuals attempt to take into account the cardinal properties of the rewards which they, as a set, stand to gain as a result of any distributive principle. Moreover, other experimenters have found other constraints enter upon group decisions of a related nature (Hoffman and Spitzer, 1984; Marwell and Ames, 1979, 1980). We conjecture, therefore, that, in a Rawlsian type original position, individuals will not choose simply to maximize the welfare of the worst-off individual. Rather, they will wish to take into account more attributes of the situation. Individuals will be concerned with the trade-offs between the floor, the ceiling, the mean, and so forth. Thus, the group decision will center around a principle which balances the trade-offs that must be made. Even in the austerity of an original position, individuals are faced with trade-offs similar to those faced in any economic decision. This means that we might expect a considerable degree of consensus in the original position, perhaps even unanimity as Rawls conjectures. But we predict a different outcome—one which is less egalitarian, yet takes into account the welfare of the poor.

Specifically, we believe that individuals placed in a Rawlsian position would choose a compound principle. Rawls referred to these compounds as “intuitionistic principle(s) of justice” and also “mixed conceptions of justice” (pp. 34–40, 316). They permit individuals to weigh a set of values flexibly and to arrive at an outcome that reflects a more complex idea of justice. Two such principles are explicitly considered (and rejected) by Rawls: (1) maximizing the average income with a constraint on the range between the best and worst off and (2) maximizing the average income with a constraint on the floor, or the income of the worst off.

To test some of these matters, the experiments we have conducted try to parallel the thought experiments posited by Rawls. Specifically, we have tried to incorporate, in a laboratory setting, a number of the features Rawls used to induce impartial reasoning via imperfect information. To the extent that the general conditions in Rawls’s original

predictions of hypotheses which are defined for that ideal must increase in accuracy or the theory becomes uninteresting. Some theories are robust in only *some* dimensions. (E.g., in neoclassical market economics substantial difficulties arise when certain violations of the idealizations occur.) Failure to exhibit “global robustness” constitutes a vulnerability or brittleness which deprives any theory of power and attractiveness. We are indebted to Norman Schofield for pointing out these relationships.

position can be relaxed to yield an empirical realization, we believe that they constitute a useful construct for deriving principles of justice. In doing so, we hope to determine whether agreement on what constitutes justice as distributive fairness, within a realizable context, has a content similar to that which Rawls postulated. But our experimental conditions also capture generic characteristics of situations inducing impartial reasoning and preferences for principles of fair distribution. We hope these permit us to determine whether an underlying consensus can be achieved about distributive principles.

Research Design and Procedures

The experiment itself consisted of three parts. First, five subjects were each introduced to principles of distributive justice and given experience choosing a principle which governed their first earnings. Second, the subjects were given a chance to discuss and then collectively to choose a particular distributive principle that governed their next payments. Finally, the subjects answered a questionnaire about themselves. At four points in the experiment, subjects were asked to rank order, according to their preferences, the principles of justice that were on the agenda and to indicate how sure they were of these rankings.

We ran a total of 44 runs of four experimental designs in three locations: the University of Manitoba (14), the University of Maryland (14), and Florida State University (16). Subjects were recruited from a wide variety of undergraduate classes.⁵ They were told they could volunteer to participate in an experiment on principles of justice, that they would be paid for such participation, and that the amount of money earned would depend upon the choices they would make in the experiment.

Some details of the experimental structure are quite critical for a proper interpretation and critique of our results. This especially holds for details regarding how subjects learned about principles of distributive justice; how subjects collectively considered the principles in discussion; how subjects reached a group decision about the principles; and how the choices of individuals were related to their pay. Consequently, a somewhat detailed discussion of these details follows.

Learning about Distributive Justice

After signing a consent form, subjects read a text introducing them to four of the distributive principles that Rawls required to be considered

⁵ Virtually no students were recruited from classes of the authors. We also avoided classes in which the students were likely to have studied Rawls and other theories of distributive justice.

in the original position: (1) the Rawlsian principle of maximizing the floor; (2) the principle of maximizing the average; (3) the principle of maximizing the average with a floor constraint; and (4) the principle of maximizing the average with a range constraint.⁶

After a short definition of each principle, subjects were asked to rank them from most to least preferred and to indicate their confidence in this ranking.

Subjects then read a text which described how each principle could lead to different income distributions being selected as the most in conformity with the particular criterion (and hence, according to this criterion, the most preferred). To illustrate, income distributions were included that, implicitly, invoked some of the trade-offs which might result from choosing one principle rather than another. Thus (see Table 1), a choice of maximizing the average with a floor constraint of \$12,000 in Table 1 would identify distribution 1 as the most conforming (and hence, most desirable) one. That distribution has a lower average income than distributions 2, 3, and 4. Thus, subjects were implicitly shown that choosing a constraint necessarily led to a compromise. They had to give up the possibility of the highest average income in order to guarantee a certain minimum level for the floor (or a certain maximum level for the range). Each subject then took a test to ensure that he or she understood the material thoroughly. Participation in the rest of the experiment was contingent upon successful completion of the test regarding the principles. Often subjects failed the test the first time, but all subjects eventually passed the tests.⁷ After successfully completing the test, subjects, once again, were asked to rank the principles.

Only after the test could the subjects earn money. Each was presented with a situation containing four income distributions with five income classes in each distribution (see Table 1 for an example). They selected a principle of distributive justice from those which had been introduced in the earlier readings. Their choice of a principle of justice governed the selection of one of the four income distributions. Then they were randomly assigned to one of the five income classes, and their payoff was determined by that random assignment. For example, suppose, when confronted with the distributions in Table 1, an individual chose to maximize the floor. This implied selection of distribution 4 (which has the highest floor). The subject then (blindly) pulled a chit from an envelope which assigned him or her to an income class. The subject was paid at a

⁶ All other distributive principles mentioned by Rawls are logically equivalent to one of these in the context of our experimental situation.

⁷ Subjects who answered any question incorrectly on the test were given some additional coaching, reread some of the material, and retook the test.

TABLE 1
Choices with Payoffs: Situation A

Income Class	Income Distribution in Dollars			
	1	2	3	4
High	28,000	35,000	30,000	25,000
Medium high	25,000	30,000	29,000	22,000
Medium	20,000	25,000	28,000	19,000
Medium low	15,000	15,000	27,000	16,000
Low	12,000	10,000	6,000	13,000
Average income	20,000	23,500	26,000	19,000
Floor or low income	12,000	10,000	6,000	13,000
Range	16,000	25,000	24,000	12,000

rate of 10 cents on each \$1,000 of yearly income for a person in that class within that income distribution.

To ensure that subjects were acquainted with the trade-offs inherent in the choice of principles, the chit drawn from the envelope specified more information than the subject's payoff. That payoff stemmed from their class assignment within the income distribution conforming to *their* choice of principle. The chit also indicated how much money they *would* have received, with that class assignment, had they chosen each of the other principles.

This gave subjects a feel for the consequences of their choices early in the experiment. More particularly, subjects could observe that no choice could result in a higher floor (average) than that obtainable by choosing to maximize the floor (average). Similarly, they could learn that maximizing the average with a constraint necessarily led to a compromise between a higher average and the item being constrained. After completing their choices, the subjects were asked to record their preference rankings for the third time.

Choosing a Principle for the Group

In the second part of the experiment, we attempted to induce impartial reasoning via imperfect information. In this part, subjects had two tasks: first, they discussed, as a group, the principles of justice. Then as a group, they attempted to adopt a principle, by voting. Subjects were told that a large number of income distributions (payoff vectors) were available for selection. They also knew

1. If they were able to agree *unanimously* on a single principle of distributive justice, that they, as a group, would receive a payoff vector from

the subset conforming to that principle. They, as individuals, would be paid off, as before, by random assignment to a class within that income distribution (payoff vector).

2. On the other hand, if they failed to achieve consensus on a principle, an income distribution from the *full* set of possibilities would be randomly chosen. They would then be paid according to it (again via a random assignment of class). Subjects were informed that the stakes in this part of the experiment greatly exceeded those of the earlier choice situations.

Discussion of the principles was to last for a minimum of five minutes; termination of this phase was complete only after the subjects unanimously agreed (both verbally and by secret ballot) to end discussion and to make a choice. In the choice phase, subjects were asked to vote over a specific agenda established by the group. A unanimous vote was required for the adoption of a principle; anything less than unanimity resulted in either a return to the discussion phase or the adoption of no principle.

In effect, subjects made their choice under conditions of extremely imperfect information. The income distributions available for payoff were not known to the subjects at any time. In addition, class positions were randomly selected after a principle was chosen. However, even then, these restrictions constituted only a very thin “veil of ignorance.” In this sense, imperfect information was used to attempt to induce impartial reasoning regarding principles of distribution.

Compensation of the Subjects

How were levels of payment determined? In considering how to pay subjects in an experiment dealing with fairness of economic distribution, some specific aspects must be considered. After all, analysts, writing about distributive justice, have argued that a variety of variables are potentially relevant: the average level of economic well-being, the depths of poverty to which an individual might fall, the upper reaches of wealth that the fortunate might acquire, some of the statistical characteristics of the distribution (e.g., the range, the standard deviation, etc.), and the notion of entitlements in general (and more particularly, deserts, or compensation for effort and socially valuable contributions). These factors had to be considered in setting up a schedule for compensating subjects.

Rawls, for example, maintained that for a sense of distributive justice to emerge, two aspects of the economic conditions under which the decisions would be made must be specified. First, he noted that the overall wealth of the society must be great enough to permit individuals to have a level of economic well-being to provide for the necessities of life. Rawls

calls this level one of “moderate scarcity.” To mirror this, we attempted to keep the average compensation at a level which approximately reflected average earned income in North America: about \$10 per hour. To further emphasize the link to subjects’ experience with moderate scarcity, the amounts in the illustrations were always expressed in terms of annual incomes, and the average amounts shown in any one set of illustrations were above the poverty level, but less than 1.25 times the average earned income in North America.

The other aspect of the economic conditions noted by Rawls was far harder to deal with in the experiment. He argued that the major motivating force for a principle of distributive justice would be the desire to stay out of poverty. Since poverty was not a possible consequence of participation in the experiment, no *direct* implementation of the Rawlsian conditions were feasible.

The Final Questionnaire

Finally, each subject filled out a questionnaire designed to provide demographic, sociological, attitudinal, and psychological data for analysis. In it, each subject also provided a final ranking of the four principles and a final indication of how sure he/she was of that ranking.

Variations in the Basic Experimental Design

The experimental objective was to determine how well imperfect information could induce impartial reasoning and thereby generate preferences and even consensus for principles of distributive justice. One obvious objection to our allusion to Rawls’s formulation as a baseline model is the question of stakes. Perhaps the stakes were not high enough to test Rawls’s contention fairly? In an attempt to address this objection, we introduced two variables which might be expected to impact subjects’ choices. The results of these variations were to be contrasted with the results in the original, or baseline, experiments to yield an indication of the robustness of the results within the range of variation available to the experimenters.

Payoffs with Higher Variance

It has been shown that an individual’s choice is often a function of the framing of the situation (Tversky and Kahneman, 1981). One way to reframe our experiment to make the stakes appear higher for the players (and thus maximize the likelihood of the selection of the difference principle) would be to alter the income distributions in part 1 so that the floors are nastier and the ceilings higher. This, in effect, raises the variance in the payoffs that subjects experienced by selecting principles.

And if rational individuals consider nothing but the floor when selecting a principle of justice, then this alternation should certainly work to Rawls's advantage. Of course, the absolute value of the payoffs would still fall far short of the "life chances" which Rawls posits for the original position, but to the extent that subjects consider the scenario seriously, the reframing should work in Rawls's favor if his arguments have merit.⁸

Choices Involving Losses: \$40 Credits

Lowering the floors and raising the ceilings in part 1 of our experiment is not the only way to reframe the situation for our subjects. Tversky and Kahneman (1981) contend that for individuals "the response to losses is more extreme than the response to gains" (p. 454). We conjectured, therefore, that if individuals were given an apparently concrete payoff before their choice of a principle of justice, their concern with a loss would lead them to focus attention on the floor. This would, presumably, increase the probability of choosing a principle which guaranteed a maximum floor.⁹

To induce this effect we changed the design slightly by giving subjects a \$40 credit prior to the second part of the experiment. In this variant they were told that their choice of a redistributive principle would yield a payoff which was a *reduction* of that \$40 credit. This, in fact, meant that subjects were to choose how to distribute losses from each of their \$40 credits. Presumably, this change raised the stakes in the eyes of the participants. But this was purely subjective. Individuals were given reason to believe that they could win \$40 in part 2 (and in fact they could), but the final returns which were available to them were identical to those available to subjects in the other variants of the experiment. This variation was run with both the original and the higher variance payoffs.¹⁰

Results

Recall that we ran a total of 44 runs of four experimental designs. Fourteen of the runs were with the original payoffs and gains, 16 experiments involved the original payoffs and losses, eight involved the payoffs

⁸ In the context of our robustness argument above, this variation should bring one closer to the ideal of Rawls's original position and yield results closer to those he posited—if his arguments are robust and the degree of variance is consequential enough.

⁹ The reader should note that the results reported below are at variance with this interpretation of Tversky and Kahneman's hypothesis. We shall examine this, and its implications, in the discussion section.

¹⁰ To make the \$40 credit credible, the amount was actually written on each subject's payoff sheet. This sheet, a cashable credit slip, was the record of the individual's earnings.

with higher variance and gains, and six were high variance with losses (see Table 2).

The Choices

Rawls hypothesizes that the unanimous choice of principle would be the difference principle (maximizing the floor). As is clear from Table 2, our results are strongly supportive of one, and strongly at variance with the other, of these claims. Under all experimental conditions, *all groups reached consensus and no group ever selected maximizing the floor as their preferred principle*. Overwhelmingly, the most popular principle was the "intuitionistic" principle of maximizing the average with a floor constraint. It was the choice of 35 of the 44 groups. This principle was chosen by all groups in our original design and was, overall, the most popular choice under all variants. Selected floor constraints ranged from a minimum of \$6,125 to a maximum of \$17,225, with a mean floor constraint of \$10,130. Seven groups chose to maximize the average, and two to maximize the average subject to a range constraint.

The Rankings

In addition to the failure of the difference principle to be chosen by any group, individual rankings solicited at the end of the experiment

TABLE 2
Groups' Choices of Principles

	Maximizing the Floor	Maximizing the Average	Maximizing the Average with a Floor Constraint	Maximizing the Average with a Range Constraint	Total
Original payoffs	0	0	14	0	14
Higher variance payoffs	0	1	6	1	8
Losses	0	5	11	0	16
Higher variance losses	0	1	4	1	6
Total	0	7	35	2	

demonstrate the weakness of maximizing the floor vis-à-vis the other principles (Table 3). Out of a possible total of 220, *maximizing the floor has the lowest number of first-place rankings ($N = 9$) and the highest number of last-place rankings ($N = 106$)*. Maximizing the average with a floor constraint received more first-place rankings ($N = 150$) than all other principles. It also has, by far, the lowest number of last-place rankings ($N = 3$). Maximizing the average is the principle with the second highest number of first-place rankings ($N = 48$).

Is this overwhelming preference for maximizing the average with a floor constraint a product of the experimental design, a reflection of a set of antecedent preferences, or a combination of the two? Evidence suggests the latter: both factors play a role. A measure of support for each principle was created by weighting the preference rankings reported by subjects (3 for first place, 2 for second place, 1 for third place, and 0 for fourth place). The average support for each principle of justice at each phase of the experiment is reported in Table 4.

As can be seen from the table, there were few significant changes in the overall support for the various principles. The aggregate preferences were relatively stable over time. For example, maximizing the floor lost support from the third to the fourth ranking (i.e., during the group discussion period), but the drop is not statistically significant. However, the gain in support for maximizing the average with a floor constraint is significant over the same time periods (t -value = -4.28 , $p < .001$). In addition, the drops in support for maximizing the average with a range constraint between the second and third rankings and the third and fourth rankings are also both statistically significant (t -value = 2.77 , $p = .006$; t -value = 2.73 , $p = .007$, respectively). The increase in support for maximizing the average between the second and third rankings is also statistically significant (t -value = -2.57 , $p = .011$). It appears that

TABLE 3
Individuals' Rankings of Principles at the End of Each Experiment

Rankings	Maximizing the Floor	Maximizing the Average	Maximizing the Average with a Floor Constraint	Maximizing the Average with a Range Constraint
1st place	9	48	150	12
2nd place	30	91	52	46
3rd place	74	47	14	84
4th place	106	33	3	77

TABLE 4
Average Support for the Principles of Justice

Rankings	Maximizing the Floor	Maximizing the Average	Maximizing the Average with a Floor Constraint	Maximizing the Average with a Range Constraint
1st place	.85	1.58	2.42	1.15
2nd place	.81	1.58	2.39	1.23
3rd place	.80	1.73	2.39	1.09
4th place	.73	1.72	2.59	0.96

the discussion behind the “veil of ignorance” changed people’s mind: it increased the popularity of maximizing the average with a floor constraint, while the principle of maximizing the average gained favor during the choice situations in part 1 of the experiment.

Another measure of the impact of the experimental procedure is given by the net number of shifts into and out of first place in individuals’ preference rankings as the experiment progressed. The data on the net number of switches into first place for each of the principles of justice are given in Table 5. Maximizing the average with a floor constraint, the most popular principle at the outset, gained still more first-place support over the course of the experiment. The largest jump in popularity occurred as subjects attempted to reach agreement on a principle of justice. Then, maximizing the average with a floor constraint gained support at the expense of the other principles.

TABLE 5
Net First-Place Gains of Principles of Justice between Records

Rankings	Maximizing the Floor	Maximizing the Average	Maximizing the Average with a Floor Constraint	Maximizing the Average with a Range Constraint
1st to 2nd record	0	5	−5	0
2nd to 3rd record	−6	8	3	−5
3rd to 4th record	−6	−8	26	−12

Subjects were also asked to indicate, on a five-point scale, how confident they were of their rankings (5 = very sure, 1 = very unsure). The average assuredness attached to these rankings increased as the experiment wore on. At the time of the subjects' first ranking, the average score was 3.56. By the end of the experiment, the subjects' average level of assuredness was 4.01. The t -value for this difference is -6.68 with an associated probability of $p < .001$. Subjects apparently became more confident regarding their rankings of the principles as a result of the experimental procedures.

The Effects of Variations in Experimental Variables

As indicated, two experimental variables, higher variance payoffs and choices involving losses rather than gains, were introduced to raise the subjective stakes in the experiments. It was anticipated that raising the stakes in this manner would change the incentive structure and make the "floor" a more threatening, and hence more salient, element in subjects' decisions.

Original or Baseline Experiment

In these experiments *all 14 groups reached consensus on a single principle: maximizing the average with a floor constraint*. Indeed, it was the unexpectedly unanimous support for this mixed, or "intuitionistic," principle that prompted us to introduce the control variables as an attempt to determine the robustness of the result.

Higher Variance Payoffs Experiment

The minimum incomes achievable for the situations in part 1 of this experiment were considerably lower than in the original experiment. But as Table 2 indicates, groups in this experiment also often selected maximizing the average with a floor constraint and never selected maximizing the floor. Of the eight higher variance payoff experiments, seven groups selected an "intuitionistic" principle (six chose maximizing the average with a floor constraint, and one chose maximizing the average with a range constraint) while one group elected to maximize the average.

Experiments Involving Losses: Original Payoffs

In these experiments, individuals were given a \$40 credit, and the group's task was to decide on a principle to govern the distribution of *losses* from their \$40 credits. The experimental results in this type of experiment are a bit different from those of the original experiment. Although unanimous agreement was always reached, out of 16 experiments, maximizing the average was selected five times. (Maximizing the

average with a floor constraint was selected 11 times.) These results do not support the interpretation of Tversky and Kahneman offered above.¹¹ None of the groups, when given a situation with increased salience of losses, increased their focus on the floor in any way. Maximizing the floor was never selected. Rather, several selected a principle which maximized the average income.

Experiments Involving Losses: Higher Variance Payoffs

These experiments were simply a combination of the higher variance payoffs and the distribution of losses. Again, the intention was simply to increase the stakes as a means of giving the subjects higher incentives to choose the difference principle. But again, the overwhelming preference of the groups was for maximizing the average with a floor constraint. Of the six groups, four chose this principle, while one chose to maximize the average with a range constraint and one to maximize the average.

Analysis

What impact did our attempts to raise the stakes, by increasing the variance in the payoffs and introducing losses rather than gains, have on the choices of our subjects? In all variants of our experiments, maximizing the average with a floor constraint was (by far) the most popular choice. Nevertheless, in the two sorts of loss experiments and in the higher variance in payoff experiments, a higher proportion of the group choices were to maximize the average (and a few choices were to maximize the average with a range constraint). This was reflected in the choices made by groups, in the rankings of individuals, and also in the *changes* in the individual rankings. Table 6 contains the results of an analysis of the changes in average rankings of the principles. In three of the four experimental conditions, the average ranking of maximizing the average with a floor constraint (MAFC) increased significantly during the discussion phase. But as can be seen from the table, the preference shifts toward MAFC were insignificant in the high-variance loss experiments. Furthermore, in the loss experiments, a second aspect of the preference shifts of the subjects are significantly different from those of the subjects in the gains experiments. The support of individuals for the principle of maximizing the average in loss distribution experiments increased significantly after receiving the \$40 credit and engaging in group discussion. Subjects in the experimental groups in which gains were distributed showed no significant change in their rankings of that principle over the period of the group discussion.

¹¹ This is discussed in the next section.

TABLE 6
Changes in Average Rankings of Principles

Type of Experiments	Average Ranking	<i>t</i> -value	<i>p</i>
Original payoffs			
Maximizing the average			
Rankings (1 + 2)	1.40	0.30	.764
Ranking 4	1.37		
Maximizing the average with a floor constraint			
Rankings (1 + 2)	2.59	-3.59	.001
Ranking 4	2.83		
Original payoffs—losses			
Maximizing the average			
Rankings (1 + 2)	1.83	-2.22	.029
Ranking 4	2.02		
Maximizing the average with a floor constraint			
Rankings (1 + 2)	2.30	-2.87	.005
Ranking 4	2.55		
Higher variance payoffs			
Maximizing the average			
Rankings (1 + 2)	1.59	-0.48	.635
Ranking 4	1.65		
Maximizing the average with a floor constraint			
Rankings (1 + 2)	2.30	-2.33	.025
Ranking 4	2.46		
Higher variance with losses			
Maximizing the average			
Rankings (1 + 2)	1.34	-2.33	.028
Ranking 4	1.79		
Maximizing the average with a floor constraint			
Rankings (1 + 2)	2.36	0.39	.698
Ranking 4	2.31		

Before one can draw any conclusions, however, regarding the effect of experimental conditions an anomaly which occurred in the loss experiments must be reported. The individuals recruited into those experiments showed, on the average, a higher *initial* support level for maximizing the average than did subjects in the gain experiments. It might be thought that the initial popularity of that principle led to the growing popularity of maximizing the average in the loss experiments. That is, perhaps the increasing ranking of that principle over the course of the loss experiments was attributable to group pressure from the individuals who had stronger feelings about the desirability of that principle in those groups. Were this the case, the increased popularity of maximizing the average in the loss experiments would be an artifact of recruitment rather than an effect of the experimental variable "losses."

As a test of this hypothesis, an analysis was performed which broke all subjects into two groups: those with low initial support for maximizing the average (i.e., ranked it third or lower) and those with a high initial support (i.e., ranked it second or better). These two groups were further broken up into individuals in gain versus loss experiments. The increase in support for the principle of maximizing the average over the course of the experiment (between the first and last measure of preference) for the individuals in the four groups was computed and compared. Table 7 gives the results of this comparison.

By "controlling" for high and low initial rankings of maximizing the average, the impact of the two types of experiments is apparent. Among

TABLE 7
Changes in Average Ranking of Maximizing the Average

Type of Experiment	Change in Average Ranking	<i>t</i> -value	<i>p</i>
Individuals with low initial support for maximizing the average			
Gains: Rankings (4–1)	4.21	1.92	.058
Losses: Rankings (4–1)	7.39		
Individuals with high initial support for maximizing the average			
Gains: Rankings (4–1)	–4.53	2.07	.041
Losses: Rankings (4–1)	–1.40		

individuals with low initial support for the principle, support increases over the course of the experiment, but it is more pronounced among the individuals in the loss experiments. The difference is close to significant at the .05 level ($t = 1.92$, $p = .058$). Similarly among individuals with high initial levels of support, there is a decrease in support over the course of the experiment, with a smaller decrease in the loss experiments ($t = 2.07$, $p = .041$). Thus, the loss experiments appear to make the principle of maximizing the average more attractive than do the gain experiments. There was no corresponding difference in the attractiveness of maximizing the average with a floor constraint between the two types of experiments. This coupled with a higher initial average ranking of maximizing the average among participants in the loss experiments would appear to explain the higher incidence of group choices of that principle in the loss experiments.

It is also worth noting that groups selecting to maximize the average with a floor constraint in the loss distribution experiment demonstrated a somewhat greater willingness to take risks to preserve their \$40 payoff than participants in the other experiments. Their mean floor constraint (in the loss distribution experiments) was \$9,439 compared to a mean floor constraint of \$10,676 in the gain experiments.¹² However, this difference is not significant at the .10 level ($t = -1.52$).

If the experimental variations had some impact on the subjects' choices, it was *not* to enhance the standing of the difference principle. Any gains in standing went to the maximize the average principle. In all cases, the principle which was the least preferred remained the maximin principle identified by Rawls.

Admittedly, our stakes were very small in relation to those which Rawls posits. But even these stakes were enough to lead our subjects to agree that the stakes were important in their decision making. In response to a question which explicitly asks whether the monetary stakes in the experiment were sufficient to affect their choices, 47 percent indicated that they either agreed or agreed strongly with the statement. Only 30 percent felt that the monetary award was inadequate. The tapes of their discussions further bear this out.¹³ In fact, the money is considerable and they know it: in part 1 they averaged \$8 earnings, and in part 2 we stressed that the stakes were substantially higher than in part 1. Indeed, in part 2 many subjects were reaching decisions while holding \$40 in hand.

¹² Excluded from our calculation of this average is one group which specified a floor to be \$2,000 below the highest minimum attainable, rather than in actual dollar amounts.

¹³ All discussions have been taped and are available for analysis.

Other Factors Affecting Preferences

There is a second major area in which our experiments deviated from Rawls's ideal: we could banish from the subjects' minds neither knowledge of their own tastes, skills, social position, aspirations, nor knowledge of the society in which they currently live. That is, we could not replicate the Rawlsian information conditions, or veil of ignorance. We could only arrange the payoffs from the experiment so that individuals would not know the sorts of income distributions from which they were choosing and in which income (payoff) class they would sit at the end of the experiment.

Therefore, the question remains: How does the very permeability of the "veil of ignorance" affect subjects' choices? To answer this, we measured such factors as risk aversion, economic status, income aspirations, and political ideology. By correlating the individuals' responses to select questions with their preferences regarding principles of justice, we can determine if any of these factors had a significant impact on the resultant choices.

Risk Aversion

Rawls and Harsanyi both indicate that individuals' attitudes regarding risk might be a factor in their selection of a principle. We included two questions to measure individuals' risk propensities. Subjects were asked to give maximum purchase and minimum selling prices for two specified lotteries with explicit odds and monetary payoffs. Their risk propensity was assumed to be directly and positively related to the prices which they quoted. An index of risk propensity was then constructed.¹⁴ The average risk propensities of the individuals choosing to maximize the average with a floor constraint was compared to the average of those who chose to maximize the average. No significant difference was found between the two groups.¹⁵

Economic Status

The subjects' socioeconomic status might be expected to affect preferences for the principles. We had a varied set of indicators of the

¹⁴Specifically, the index consisted of the ratio of the buying price to one lottery's mathematical expectation plus the ratio of the selling price to the mathematical expectation of the other ticket. This ratio was normalized to make a value of 1 equivalent to risk neutrality.

¹⁵As a further check, individuals' scores on the index were correlated with individuals' average support for the four principles. Only one weak correlation of .107, significant at the .07 level, was found, and it was in an unexpected direction. Increasing risk propensity was positively correlated with support for Rawls's principle.

subjects' socioeconomic status. These included how many hours they were working while going to college and what percentage of their college expenses were met by parents, by their own income, and by loans. As before, the sample was broken into two groups: those who chose to maximize the average and those who chose to maximize the average with a floor constraint. Mean scores of members of the two groups were compared to determine whether there were significant differences between them.

Overall, none of the differences were highly significant. However, as is apparent from Table 8, three of the variables approached significance. Differences in means significant at the .10 level (but not at the .05 level) were found in the number of hours per week worked while going to college, the percentage of college expenses met by their own income, and the percentage of college expenses met by parents.

These results seem to indicate that increasing socioeconomic status is negatively related to support for the floor constraint principle. Those who choose to maximize the average, worked fewer hours, used less of their own income to meet college expenses, and relied more heavily on their parents than individuals selecting the floor constraint. An ancillary test of the impact of these variables was performed by correlating them against average support for the various indices. Not too surprisingly, the percentage of college expenses met by parents was negatively correlated with average levels of support for the floor constraint principle (-0.155 correlation with $p = .013$). In a similar vein, positive (but weak) correlations were found between support for the floor constraint and the percentage of a subject's college education financed either by loans or the student's own income ($r = .134$, $.124$ and with $p = .028$, $.038$, respectively).¹⁶ Higher levels of parental support may well have a negative influence on their concern regarding the floor income. To the extent that this is true, it points to the permeability of our "veil of ignorance."

Income Aspiration

Another factor which might penetrate our veil of ignorance is the subjects' income aspirations. To tap the relationship between income aspirations and support for the principles, we needed a measure of aspirations. Subjects were asked to name their minimum acceptable salary level at three different points in the future: at their first job, at age 35, and at

¹⁶ The only other significant result was an unexpected negative correlation between support for the maximizing the average and the percentage of college expenses met by trust moneys ($r = -.125$, $p = .037$).

TABLE 8
Differences in Three Socioeconomic Indicators between
Subjects Choosing Different Principles

Type of Experiments	Mean	<i>t</i> -value	<i>p</i>
Percentage of college expenses met by parents			
Maximize average (<i>n</i> = 35)	57.14		
Maximize average with floor constraint (<i>n</i> = 168)	42.64	-1.95	.053
Percentage of college expenses met by own income			
Maximize average (<i>n</i> = 35)	22.80		
Maximize average with floor constraint (<i>n</i> = 169)	35.69	1.86	.065
Number of hours worked			
Maximize average (<i>n</i> = 35)	4.31		
Maximize average with floor constraint (<i>n</i> = 173)	7.50	1.76	.08

age 50. Responses were aggregated into an index. The scores of individuals in the two groups were compared to determine if there were significant differences which could explain the different choices. There were no significant differences in the mean income aspirations of those who chose to maximize the average with a floor constraint and those who chose to maximize the average.

Political Ideology

Another factor we conjectured would affect a subject's choices of a principle was his or her underlying political ideology. All subjects were asked to rank themselves on a five-point conservatism (1)–liberalism (5) scale. These scores were then related to their choices of principles.

There were no significant differences in the mean ideological scores of those who chose to maximize the average compared to those who preferred a floor constraint. As a further test, we correlated the individual's ideological scores with his or her mean support for the four principles of justice. The floor constraint principle was uncorrelated with the variable in question ($r = -.013$, $p = .426$). However, liberalism was positively (but weakly) correlated with support for maximizing the floor ($r = .148$, $p = .016$) and negatively correlated with support for maximizing the average ($r = -.184$, $p = .004$).¹⁷

Discussion

Major Results

Our experiments were designed as empirical tests of three main aspects of theories of justice as fair division. First, could imperfect information induce preferences for principles of distributive justice so strongly that unanimous agreements could be reached in empirical contexts? Second, would the content of that agreement be consistent? And finally, would Rawls's difference principle be chosen, or would some other principle emerge? Quite obviously, we also wished to establish the feasibility and value of subjecting aspects of ethical theories to empirical testing.

With regard to the use of imperfect information, our results are positive and unambiguous. In *every* case the introduction of imperfect information seems to have harnessed the self-interest of the subjects and allowed them to reach unanimous consensus on a principle of distributive justice. To that extent our experiments bear out the predictions of both Rawls and Harsanyi. It appears that a "veil of ignorance" permits a group to reach stable decisions in which they have considerable confidence.

But another of our results strikes at the heart of both Harsanyi's and Rawls's substantive arguments: the content of the agreements. Overwhelmingly, the most popular principle for economic distribution was not the candidates they proposed but rather the "intuitionistic" principle of maximizing the average with a floor constraint. It was chosen most often in all variants of the experiments. Rawls's difference principle was *never* chosen and was the least popular in preference rankings. Harsanyi's principle of maximizing the average was a very distant second in popularity as both a group and an individual choice.

¹⁷ It may be that Canadian and American subjects differed in their interpretation of the connotations of the words "liberalism" and "conservatism." Nevertheless, the general direction of the relationship appears not to be affected by these differences in nuance.

The intuitionistic principle was the clear winner. This result relates to the psychology of individual choice. When considering one principle which concentrated *only* on the floor and one that focused *only* on the average, the subjects preferred a compromise. This implies that *individuals treat choices between principles as involving marginal decisions. Principles are much like economic goods inasmuch as individuals are willing to trade off between them.* This is an extension of results reported in Frohlich and Oppenheimer (1984) and Bond (1983). The importance of this finding goes to the core of any attempt to identify a single factor principle of justice. Our results indicate that any such search is likely to conflict with individuals' desires to take into account more than one factor in such value judgments. Nevertheless, we are not in a position to offer a precise substitute for the single factor principles. While Rawls's and Harsanyi's principles are specific solutions, our result offers a (potentially infinite) family of solutions because of individuals' latitude to set the value of the floor.¹⁸ While this may be viewed, by some, as a weakness, it accommodates a wider range of values than do the other principles. Indeed, it is encouraging to note that one might consider the intuitionistic principle a compromise, or mixture, of the Rawlsian and Harsanyian principles.

The Dynamics of the Group Decision Process

As has been mentioned, subjects were told to consider the arguments and principles in terms of annual incomes, even though the stakes were only a very small (although proportionate) fraction of an annual income. The specific information imperfection introduced experimentally simply hid the properties of the distribution of payoffs available to subjects. Nevertheless, the content of the discussions which ensued in part 2 and the seriousness with which subjects pursued the issues reinforces the notion that subjects came to grips with the normative issues with which we were concerned. A description of the "typical" dynamics of an experiment may provide insight into the nature of the decision process which took place.¹⁹

How did groups arrive at their choice of principle? Two main issues arose in virtually all discussions. Subjects were concerned that (1) no one should be reduced to insufferable poverty and (2) any redistribution rule not preclude others from *earning* higher income. This latter concern was usually argued in two different ways. First, it is not fair to take

¹⁸ We are indebted to Oran Young for this observation.

¹⁹ As noted above, the discussions were taped and are available. We have not conducted a systematic content analysis of the discussions or of their dynamics.

too much away from those who *earn* higher incomes. And second, too much redistribution might unduly limit incentives and reduce overall production.

The tension between these concerns was then the focus of much of the discussion. Groups spent *considerable* time discussing the trade-offs between setting higher constraints on the floor (and thus lowering their total income) and setting lower constraints on the floor (and thus hurting those in the lower class). Most individuals wanted to balance the security of a higher floor with the possibility of increasing the average income in the hope that they might fall into one of the higher income classes. From the comments made during the discussions, they did not want to cut too much from an attractive ceiling to obtain a higher floor. Limiting the depths to which anyone could plunge by setting an acceptable floor constraint and, at the same time, letting any surplus funds remain unredistributed, was a natural way of resolving this tension. Typically, after realizing this possibility, the group discussion turned to specifying the appropriate level for the floor. Discussions then focused on what constituted a bearable standard of living and what levels of support would be unduly unfair for higher producers to fund.

Maximizing the average with a floor constraint usually provided an acceptable means of reconciling the efficiency implicit in the principle of maximizing the average with a concern for the welfare of the poorest. As each experiment wore on, more subjects adopted this attitude, and their confidence in their ranking of principles increased. Indeed, they appeared to be moving toward an equilibrium between the group's choice and their own notions regarding fairness.

Moreover, during the discussions, subjects explicitly differentiated between the principles in various ways. First, as indicated above, most subjects wanted a chance to make an unconstrained amount more than the minimum. This led to serious discussions as to what constituted a "proper" minimum and led to considerable variation in the floors which were adopted. The minimum floor constraint selected was \$6,126, and the maximum floor was \$17,225 with a mean floor constraint of \$10,130. In experiments distributing losses, when the subjects felt they had a "high" possible ceiling to "protect," the floor constraints selected were *lowered*. The spreads in the selection of floor constraints lead us to question Rawls's idea that any selected floor constraint would be so high as actually to equal the maximizing floor principle. The observed social dynamics in the experiment, and the numbers themselves, lead us to believe that this was not what went on.

The intuitionistic principles win out, we would argue, not because of subjects' lack of knowledge or understanding, but because they

reflect the complex realities behind the judgments. Individuals have multiple concerns regarding income distributions. These concerns are not all reducible to a single-stranded principle. Discussions in the experiments show that there is a consensus that “everyone be lifted beyond poverty.” Individuals also agree that one must think about the floor constraint in terms of what it causes one to give up. Beyond this, there is little agreement. Some would lift individuals to a “position where they can enjoy self-respect”; others would ensure that only the bare necessities be available. What Rawls failed to take seriously were the potential indeterminacies in individual values. Even in Rawls’s pristine original position, with individuals quite deprived of information, he must allow them *some* residual values. Without these values, there is no basis for ranking alternatives and, hence, no basis for rational choice. In our experiments we have demonstrated that some of these values are bound to conflict. The conflict results in trade-offs of the values which are at stake in the decision.

On the Experiment’s Distance from Ideal Conditions

As admitted above, our experimental context falls far short of Rawls’s idealized conditions. Therefore, we introduced two experimental variables to discover the possible disturbing effects of the stakes and also gathered background information on the subjects to check for the impact of the transparency of our veil of ignorance. Regarding the latter, we found no *major* effects of such factors as risk propensity, political ideology, salary aspirations, and socioeconomic status on the subjects’ choices of principles of distributive justice. Although a few significant differences were found with some of the socioeconomic status variables, they clearly were not strong enough to account for the overwhelming unpopularity of Rawls’s principle.

As to the impact of the stakes, the reader could argue that, since they were relatively small, the experiments were too remote from ideal conditions to be fair tests of any theories of distributive justice as fairness (although \$40 for the second hour of the experiment is not an insignificant sum). Most certainly, there is a wide gulf between our payoffs and lifetime earnings. But too much emphasis on that difference tends to obscure a major analytic point in support of the results we obtained. Rawls’s *primary* argument for maximizing the floor is the horrible downside risk of poverty. *That risk can be avoided precisely to the extent desired by the subjects. They need only choose a particular floor constraint.* Moreover, this frees them to consider other values (such as increasing average income) in their decision. It could be that in a Rawlsian original position, with its potential for great vagaries,

different conceptions of the appropriate floor constraint could emerge. But, given our results, it would appear most likely that individuals would find a role for *some* floor constraint. Moreover, it would be surprising if they set that floor constraint so high that it would be the *maximum* attainable.

Although we cannot know whether our results would appear robust as one approached Rawls's conditions more closely, the results from our different experimental conditions gives us some clear clues.

The Rawls-Harsanyi approach to inducing impartiality and consensus on fairness seems to work consistently, even in our quite far-from-perfect laboratory realizations of their models: hence, this conclusion of their models seems quite robust. On the other hand, there is *no* support for the Rawlsian conclusion that the choice of principle will be to maximize the floor. It is unpopular and appears wrong, or, at least, not robust.

As to the robustness of our own positive findings regarding the choice of maximizing the average with a floor constraint, we found considerable support for it under all test conditions. But the support is not completely homogeneous. In the distribution of loss experiments, support for maximizing the average increased significantly (see Tables 2 and 6). Further examination of Table 6 shows that when (and only when) the experiment was restructured to distribute losses (with higher variance), support for the floor constraint principle did *not* increase during the discussion phase.

Indeed, Harsanyi's principle of maximizing the average was the second most popular rule (not floor maximization). This principle was especially appealing to groups in the distribution of losses experiment. In those variants, subjects had each been awarded \$40. A number of those groups chose as if they had much to lose by the adoption of a rule which decreased the *ceiling*. This could be attributed to subjects' added concern with the *costs* of possible losses. It would partially explain the increased popularity of a rule to maximize the average, which tended to increase the ceiling.²⁰

To see why, in the loss experiments, one could expect increased concern for a high ceiling, consider, once again, Tversky and Kahneman's arguments. They argued individuals experience high and decreasing marginal valuations for *losses* from the status quo and by comparison, smaller and decreasing marginal valuations for *gains* from a status quo. They

²⁰ Alternative explanations of these different group choices are possible. One possibility is that we had a different distribution of initial preferences in the subjects in the loss experiments than in the gain experiments.

maintain that there is an asymmetry in the way losses and gains are experienced. This has a number of implications for our experimental design, analysis, and conclusions.

For Tversky and Kahneman the definition of the status quo is an important element in one's decision calculus. This follows from the change in marginal valuation which they claim takes place at that point. The status quo is an inflection point in the valuation curve. From the status quo the individual is unwilling to take fair gambles for gain (due to decreasing marginal valuation). Moreover, he/she is also unwilling to take small sure losses to ensure against a fair gamble involving a potentially larger loss. These properties of the valuation curve imply that setting the status quo is crucial. There should be a difference between loss and gain experiments. In the latter, individuals should focus on losses, but with quite a different outcome than that expected by Rawls. Indeed, if Tversky and Kahneman are correct, individuals should be *less* interested in ensuring a high floor when losses from the status quo are likely.

We ran two experiments with losses: with the original payoffs and with higher variance payoffs. In both cases there were indications of significant increases of support for the maximizing of the average principle. In the high-variance experiment there was a substantial increase in the proportion of times the group chose this principle. Moreover, only in the low-variance loss experiment did the floor constraint principle fail to increase its support during the discussion and choice phase of the experiment (see Table 6). If we assume that the introduction of potential losses induced a new status quo, we can interpret these results as confirmation of Tversky and Kahneman's predictions. The principle chosen was somewhat sensitive to the framing.

Perhaps our findings show that some changes in preferences can be induced by reframing (as in the loss experiments). The status quo *does* seem to make a difference. The result of any impartial reasoning process appears to be a function (even if only weakly) of the status quo. Apart from marginal implications for our analysis, this interpretation of the results has broader implications for theories of distributive justice based on imperfect information. For Rawls and others, the framing of the decision problem, including the status quo, is a crucial determinant of the ultimate decision. Yet Rawls's (very thick) veil of ignorance can be seen to prevent individuals from identifying a status quo with any precision.

If the definition of a status quo, or the absence of one, frames the situation and "drives" the result, then no decision context can be *ethically neutral*. As Tversky and Kahneman state at the conclusion of their article: "When framing influences the experience of consequences, the adoption of a decision frame is an ethically significant act" (p. 458).

Testing Ethical Theories

We have tested the viability of procedures which use imperfect information to induce preferences for fair distributions. The experiments have been developed in relation to John Rawls's formulation. This particular reference should not obscure the broader argument that is implicit in the procedures we have adopted. If these results have any merit at all, it must be the case that it is meaningful and reasonable to subject some aspects of ethical theories to empirical tests.

Whenever an ethical theory implicitly contains a model of human psychology from which normative conclusions are derived, that model opens the possibility for testing. If an ethical argument is structured with the presumption that its force can persuade a reasonable individual (however defined) of its claims, then it is necessary to specify what constitutes the psychology of a "reasonable" individual.

Where a complete specification is not provided, it becomes important to use empirical means for testing reactions via real individuals. Although this technique clearly introduces a new set of difficulties of its own (e.g., the problems of approximating ideal types, as discussed above), if used properly it casts additional light in otherwise rather dark corners. Empirical testing of ethical arguments thus may interact constructively with deductive reasoning in much the same way that thought experiments (e.g., those of Rawls) do.

Moreover, if the force of an ethical argument derives from the presumed consequences of its rules, then that force is *only* compelling if the consequences do indeed follow in the real world. Here too there is a role for testing.

The Policy Relevance of the Experiments

There is an additional argument for engaging in experiments of this sort. Our experiments demonstrated the consistent performance of a procedure and an overwhelming support for one principle of distributive justice. A principle which maximizes the average income while ensuring a tolerable floor income is broadly popular. This support crosses the entire spectrum of subjects in the sample. It appears to be a near-consensus position at the start of the experiment, and the experimental veil of ignorance only reinforces its popularity and the certainty with which individuals approve it. Given its overwhelming popularity, one would expect this principle to be apparent in the political platforms of the major political parties of Canada and the United States. And indeed in retrospect one can see it. Both countries have established elaborate sets of programs whose professed goals are to ensure that no one in

the society falls below an acceptable level of subsistence. As we know, the reality of these programs is far removed from this rhetoric.²¹

But how applicable is this to the "real world"? Where is the "veil of ignorance"? Do not we all know who we are, where we sit, and where we come from? Given this, how can we claim policy relevance for our results? One possibility is to consider the future. We may all know where we have been, but none of us can know with certainty where we are going. Thus, voters can be thought of as sitting behind a (partial, or "thin") veil of ignorance: the future is not known to them. To the extent that the real world has a "thin veil" similar to the one in our experiments, our results are relevant. Moreover, given that we can identify a near-consensus position from an ethically supportable position, the rule of maximizing the average with a floor constraint has much to commend it.

For these reasons we are drawn to the conclusion that there is a role for the experimental testing of aspects of ethical theory. These experiments are a tentative step in that direction.

Manuscript submitted 21 October 1985

Final manuscript received 14 July 1986

REFERENCES

- Bond, Pat. 1983. Game choice experiments and social interaction: Testing a simple model. Presented at the annual meeting of the Public Choice Society, Savannah, GA.
- Eavey, Cheryl L., and Gary J. Miller. 1984. Fairness in majority rule games with core. *American Journal of Political Science*, 28:570-86.
- Frohlich, Norman, and Joe A. Oppenheimer, with Irvin Boschman and Pat Bond. 1984. Beyond economic man. *Journal of Conflict Resolution*, 28:3-4.
- Frohlich, Norman, Joe A. Oppenheimer, and Cheryl Eavey. Forthcoming. Laboratory results on Rawls' principle of distributive justice. *British Journal of Political Science*.
- Harsanyi, John C. 1953. Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy*, 61:434-35.
- . 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63:309-21.
- Hochschild, Jennifer L. 1981. *What's fair: American beliefs about distributive justice*. Cambridge: Harvard University Press.
- Hoffman, Elizabeth, and Matthew L. Spitzer. 1984. Entitlements, rights, and fairness: An experimental examination of subjects' concepts of distributive justice. Mimeo., University of Maryland.
- Marwell, Gerald, and Ruth E. Ames. 1979, 1980. Experiments and the logic of collective action, I and II. *American Journal of Sociology*, 84:1335-60, 85:926-37.

²¹Page (1983) and Oppenheimer (1979) show that very few of the transfers which occur stem from political desires to improve the conditions of the current poor. Rather, most of the transfers seem to be motivated by risk aversion and the ensuing demands of the currently nonpoor for insurance type programs.

- Miller, Gary J., and Joe A. Oppenheimer. 1982. Universalism in experimental committees. *American Political Science Review*, 76:561–74.
- Nozick, Robert. 1975. *Anarchy, state, and utopia*. New York: Basic Books.
- Oppenheimer, Joe A. 1979. The democratic politics of distributive justice: Theory and practice. Mimeo., University of Maryland Distinguished Scholar-Teacher Lecture.
- Page, Benjamin I. 1983. *Who gets what from government*. Berkeley: University of California Press.
- Rawls, John. 1971. *A theory of justice*. Cambridge: Harvard University Press.
- Soltan, Karol E. 1982. Empirical studies of distributive justice. *Ethics*, 92:673–91.
- Tversky, Amos, and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *Science*, 221:453–58.
- Von Neumann, John, and Oskar Morgenstern. 1944. *Theory of games and economic behavior*. Princeton: Princeton University Press.