

Empirical Approaches to Normative Theory¹

Norman Frohlich

I.H. Asper School of Business, University of Manitoba
Winnipeg, Manitoba R3T 5V4

(204) 474-6385; e-mail: frohlic@ms.umanitoba.ca

&

Joe Oppenheimer (Corresponding author)

Department of Government & Politics
University of Maryland, College Park, Maryland 20742

(301) 405 - 4136; e-mail: joppenhe@bss2.umd.edu

Abstract

Empirical methodologies and philosophy are rarely discussed together but this essay is about how simulations and experiments may be useful in the pursuit of philosophical knowledge. It focuses on how these methods can illuminate some fundamental problems in political philosophy - viewed as a branch of ethical philosophy. It is widely accepted that the way a social decision is made can affect not only the substantive content of the decision but its ethical status. We argue that certain conditions, such as impartiality, may be necessary for the achievement of social optima and for imbuing some decisions with ethical validity. Experiments and simulations may help in the discovery of both the normative and substantive implications of various decision processes and hence the content of such properties in decision making. For ethical consequentialists this would be part of a feedback loop: the substantive implications of a process would help determine the ethical status of the decision and help identify appropriate properties for social decision making structures.

1. We are very grateful for the wonderful hospitality shown to us at GATE/CNRS where the staff and economists let us enjoy our sabbatical working on this problem. Our conceptualization of the problem, and certainly also this paper, were also enriched by the very timely comments of Thomas Schwartz, Catherine Eckel, and Michael Cain. We are, of course, also lucky to have colleagues who have been helpful in our formulations of these problems over the years in conversations too numerous to recall accurately. And finally we must acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada, the National Science Foundation (grant #01523490), and the Universities of Maryland and Manitoba for their continued support.

Table of Contents

Abstract	i
Introduction	Page 1
Understanding the Values of Individuals	Page 1
A Role for Egoism	Page 2
Dictator Experiments, Egoism and the Possibility of Other-Regarding Values	Page 3
Simulations and Experiments as a Way of Getting Beyond Egoism	Page 5
Other Contexts in Which Impartiality Can be Simulated	Page 7
The Sensitivity of Simulations of Ethically Privileged Perspectives	Page 9
References	Page 9

List of Figures and Tables

Political Philosophy, Simulations, and Experiments

Norman Frohlich & Joe Oppenheimer

Societies are constantly trying to beat people into shape because they stubbornly fail to conform to some preconceived pattern of human possibility. Political theory is in this sense an empirical discipline whose hypotheses give hostages to the future, and whose experiments can be very costly. *Nagel, 1991, p. 7.*

Nagel's position has been taken to heart neither by philosophers nor normative political theorists. Both have been reluctant to embrace empirical techniques. Yet, some empirical elements must be understood if one is to tackle the philosophical questions of justice, fairness, and welfare. Here we address extending empirical techniques, such as experiments¹, into social and political philosophy as a means of clarifying the nature of reality about which we ordinarily know too little.

Introduction

The ethical status of a social decision rests on a number of factors, and philosophers differ in the weight they assign to them. For some, process is most important (e.g. Nozick, 1974); for others the outcome or consequences matter most. But there is virtual unanimity in all of modern western political philosophy that for the outcome of a social decision to be judged good, the process must positively reflect at least some of the values of the individuals involved.

Experiments allow one to examine both decision processes and individuals' values, to see how different processes affect the connection between the values and the outcomes and to determine how different processes affect the values of the individuals. We discuss these issues and reach some tentative conclusions about how some processes, specifically ones involving a form of impartial reasoning, can be invoked in the laboratory and can help us understand what constitutes good processes and ethically right decisions.

Understanding the Values of Individuals

Any focus on the welfare consequences of a social policy (and in modern democracies most such policies are evaluated in terms of welfare consequences) involves understanding the values or preferences of individuals. If the values held by individuals were either non-empirical, or as simple to measure as, say, the height of an individual, there would be less for us to concern ourselves with in this essay. But preferences appear to be both hard to measure and extremely fragile. Their instability has been discovered and documented in a set of famous, and seemingly ever expanding, set of laboratory experiments (Kahnemann, 1982; Tversky, 1988; Quattrone, 1988; Grether and Plott, 1979).

Value instability is not a question of under-specification. An individual's values need not be straightforward and simple: they can be conditional (e.g. John prefers sour dough bread to plain white bread unless he is having a peanut butter sandwich or jam with his toast in the morning). The problem is that preferences seem to vary not just with the specifics of the choice situation but rather how the situation is described, or framed. Put another way, John may have several incompatible preference structures in his head regarding a particular situation, that can be evoked by different ways of looking at that situation.

1. When we speak of experiments here we are restricting ourselves to those involving human subjects.

This creates a big problem for the ethical philosopher. If good social policy is to reflect the values of the people in the society, and if we cannot identify a unique set of values for each individual, one is forced to ask a fundamental philosophical question: “Which of the values of each of the individuals should be singled out for representation?”² And what determines which values are tapped? When do the expressions of the individual’s most egoistic come to the fore? Is that material suitable for the construction of good social policy? Are egoistic desires the basis building ethical theory?

Beyond Egoism

Only a few philosophers argue that egoism should form the basis of ethical theory (see the discussion by Feldman, 1978). A few others object to any role for egoistic values in ethical choice (e.g. Kant). But that position is clearly too extreme for most. Individual values must count for something in determining actual decisions. There must be some fundamental desires grounded in individual need that put in train individual motivation (Damasio,1994). Egoistic values must also play a role in the normative evaluation of decisions. But exactly how central they are or should be are deep questions. Nagel (1991) has tackled this problem eloquently. He points out that a primary task facing political philosophers is to understand how to design institutions which can balance the passionate and egoistic interests of individuals regarding how they desire to live their lives, with the social interest we all have in society’s impartial fairness and justice, which must, of necessity, impose some limits on egoistic desires. The egoistic passions which underlie Hobbes, Bentham, and classical economic theory must be counterbalanced by other goods.

But if social decisions are to be good ones they cannot rest solely on particularistic values. Social values, including concern for others’ welfare must also play a role in a social decision. It is common to hear hopes that we can get social policy formed beyond the shade cast by the narrowest of private interests. Morality and ethics, and hence a proper political order, go beyond the simplest notions of egoism because ethics must also deal with how people want *others* to govern their behavior so that they themselves can be protected at the times when they might be vulnerable (Gert, 1998, p. 10). And since it is taken as a given that we cannot expect others to do what we would not do ourselves, ethical choices, choices consistent with what we would want others to do, require a modification of simple egoism. The values inside the heads of individuals must furnish the building material for the theory of moral philosophers but they must be more complex than mere egoism. Audi (1997) explicitly shows why ethical considerations must be understood to be, in part, internal to the individual.

Dictator Experiments, Egoism and the Possibility of Other-Regarding Values

Economists’ choice models are based on the assumption that individual values are egoistic: individuals do *not* have any direct concern for the welfare of others. More complex (and nicer) assumptions about human nature tend to make the mathematical models used by the discipline messy - or even intractable. As a result, economists have devised the dictator and other experiments to test for the accuracy of the egoism assumption.

2. In addition, of course, one must try to comprehend which possible political structures would be responsive to those types of values and which social environments are conducive to the generating of those values. These are questions to which we will return below.

The dictator experiment (developed by Hoffman, et. al. 1994) is used to identify the degree to which individual choices in groups could be attributed to egoism. A typical dictator experiment proceeds as follows: Individuals arrive at one of two rooms, call them Room A and Room B. A monitor is randomly chosen from subjects in Room A and runs the experiment to increase the credibility of the instructions and to reduce experimenter effects. Other subjects in that room each receive an envelope with \$10 in it. They are told that they can keep any amount of the money and leave the rest in the envelope for an individual, in Room B, with whom they have been anonymously paired. They must choose, in total privacy, and with total anonymity, how much of the money to take out of the envelope to keep for themselves and how much to leave in the envelope. After the decision, the individual leaves Room A and the experiment. What is left in the envelope is delivered to the paired person in Room B and is recorded by the experimenter. The experiment is designed so that neither the experimenter nor the recipient know the identity of the individual who leaves the money in the envelope. The design which aimed at anonymity was assumed to deprive the subject of any way of getting a reputational, or other secondary gain from the choice.

Significant numbers of subjects (about 1/2) make choices which evidence concern for others and hence falsify the strong self-interest assumption (Roth, 1995; Hoffman, McCabe and Smith, 1996). The falsifying results were assumed by experimenters to flow from inadequate levels of anonymity. Hence, the experiments were refined to increase the levels of anonymity. It was found that the higher levels of anonymity in the dictator protocols improved upon a number of earlier research designs and lead to lower levels of seemingly other-regarding behavior.³ But significant levels of other regarding behavior persisted.

A few years ago Eckel and Grossman (1996, also see Grossman and Eckel, 1996) conjectured that the relatively high level of money kept in dictator experiments might be a result of the absence of any recognizable social context provided in the experiments. They argued that the lack of context led to a disconnection of the dictator from the potential recipient. They designed alternative experiments with specific contextual understandings. Money left by dictators was contributed directly to the Red Cross. They found increased levels of money left with this additional context.

These results led us to conclude that the evocation of self-interested as opposed to other regarding values is very sensitive to context, and also potentially confounded by the anonymity requirements of the dictator experiments. Subjects, dividing the money may doubt both the existence of a second room, and whether the money they leave will go to a real subject. Hence, keeping money may be a function of both self-interest and of disbelief in the experimental protocol (Frohlich, Oppenheimer and Moore, 1998). Thus, even an other-regarding subject (i.e. a non-self-interested one) might choose not to leave money. Why leave money for no good purpose? Further, the subject may interpret the experimental setting as a competitive game. In either of those contingencies, results of seemingly self-interested behavior may be over-interpreted as evidence for the self-interest assumption.

We manipulate doubts about the existence of others and the veracity of the experimental design both by giving additional evidence for the existence of other people and another room and

3. In the early 1980's two of us (Frohlich and Oppenheimer, 1984) conducted experiments similar to dictator experiments both in motivation and in some design details. Those experiments also maintained anonymity between the chooser and recipient. But the experimenter was able to identify, at a later date, how much any individual chose to leave. Thus, those experiments might have led the subject to make a choice in order to gain reputational benefits.

by constructing conditions where all subjects are in one room. A substantial proportion of what appeared to be self-interested behavior was explicable by subjects' doubts regarding either the existence of another room or the truth of the assertion that the money left would be given to subjects. Perceptions that they were involved in a "game" (as opposed to a generic social choice situation) also explained a portion of their behavior. When subjects' doubts were diminished by compressing the experiment into one room (while preserving anonymity) the amounts of money left increased significantly.

Current efforts at establishing a purely self-interested explanation for choice seem to have foundered, or at least, stalled. The degree to which self-interest accounts for observed behavior in laboratory experiments is sensitive to attributes of the choice situation, even when anonymity is insured. Experiments framed in market contexts generate choices more in conformance with self-interest than does a 'dictator' experiment or other non-market environments (see Roth, 1995, p. 282).

A few conclusions follow: First, strict egoism (or self-interest) is not accurately descriptive of human behavior. Second, the values which an individual brings to bear on a decision context are exquisitely sensitive to the context in which the decision is framed. Hence, if we are to consider ethical decisions, extreme care must be exercised in designing the context of the decision. Third, the experimental lab or simulation, furnishes both an opportunity and a challenge to the researcher.

Experiments as a Way of Getting Beyond Egoism

Knowing that non-egoistic values enter into decisions is encouraging to those who are interested in the possibility of identifying ethically good decisions. But for the reasons alluded to above the values or preferences of individuals are fragile. They are a function of the way they think about the choices at hand. Some ways of thinking about the problem may filter out the costs and benefits to others and causes the individual to choose from a particularly egoistic point of view. Other perspectives give more weight to social costs and downplay the egoistic. Considering the same alternatives in differing lights, alters the individual's choice because the alternatives come to cast different shadows in their minds. Different social institutions may more readily invoke one perspective or another; the same is true of different experimental settings. Is it possible to differentiate among the institutions to identify some as better than others in allowing individuals to make more ethically acceptable decisions?

As the quotation at the outset specifies, social experiments can be very costly: viz. the short unhappy history of the Soviet Union. It is at this point that experimental or simulation design can help to unravel the dilemma of the appropriate perspective to take on a problem. Experiments allow us to construct different social mechanisms to observe which values come to the fore in different contexts and how they lead to different outcomes. They even allow us to get a glimpse of what are the just outcomes.

We have shown (Frohlich and Oppenheimer, 1992 and replicated by Lissowski et. al. 1991; Jackson and Hill, 1995; Saijo Takahashi, and Turnbull, 1996; and Oleson, 1997) that one can simulate, in a laboratory, aspects of John Rawls' "veil of ignorance" to induce aspects of impartial reasoning among groups of subjects and hence gain information regarding preferences over principles of distributive justice. Similar constructs might be effective in other ethically problematic situations. In fact, very significant consensus was obtained in Canada, three locations in the United States, Poland, Australia and Japan regarding the preferred principle of distributive justice.

There are two aspects of impartiality which recommend it as a way of generating insight into individual and group valuations of alternatives. First, since an impartial point of view is one that takes into account the outcome for all individuals in an evenhanded manner, it has been argued by ethicists to have normative weight. Second, individuals, discussing and deciding an issue in a group, possess an advantage over a set of the same individuals cogitating alone and reaching individual decisions. As the most rudimentary common sense tells us, each individual has private information based on individual experiences not accessible to the others unless information is shared. Group discussion broadens the perspective. Ideally it should reflect the perspective of all those affected by the decisions. But discussion in an impartial context does not only result in the increase in a subject's information content. The context may evoke empathy and understanding of how others are similar to oneself and how their needs may exercise a moral claim which could and perhaps should limit one's own egoistic desires. The impartial context may be particularly privileged in sensitizing subjects to the claims of all members of the affected group and may furnish a mechanism for integrating individual egoistic values with both the needs of others and with other more gentle sentiments of one's own.

Group discussion in a structured environment is a way of bringing to bear wider information and a variety of perspectives so that individuals can expand their understanding of the domain of choice and refine their evaluations of the rankings of different alternatives in that set. If the ethical decision is sufficiently well specified and if the sharing of information under impartiality yields a preferred perspective, then a unanimous decision based on a common understanding of the common domain (as it applies to that situation) may be obtainable.

As Thomas Nagel (1991, p.8) notes: "If we cannot, through moral theory and institutional design, reconcile an impartial concern for everyone with a view of how each individual can reasonably be expected to live, then we cannot hope to defend the general acceptability of any political order." The robustness of group choices of principles of distributive justice in a laboratory simulation of Rawls' "veil of ignorance" (albeit a highly imperfect approximation of some of the relevant aspects) give one the hope that such a reconciliation is possible.

Other Contexts in Which Impartiality Can be Simulated

Aside from issues of social justice and fairness, one of the most widely encountered and widely discussed set of ethically problematic situations are those known as social dilemmas or collective action problems.

A collective action problem can be defined generically as follows: A group of individuals faces a decision (or set of decisions) which has implications for all members of the group. The decision has the characteristic that if each individual chooses an action on the basis of a rational and self-interested calculation, the outcome for the group is sub-optimal - one that is worse than could have been achieved had some other choices been pursued. If each pursues his or her private interest in a rational and selfish fashion, the result is likely to be disastrous for the group. A group could do better and that it would be the ethically correct thing for group members to cooperate in achieving a better outcome for the group.

The "ethically correct solution" to the well known collective action problem characterized as the Prisoner's Dilemma (PD) can be deduced by applying impartial reasoning analytically to the situation facing individuals in such a situation (Frohlich, 1992, Frohlich and Oppenheimer, 1996). If one imagines playing a collective action type game from the perspective of not knowing which player's

payoffs you are going to get in the end you have to be concerned about how all the players will do as a result of your choice - since you may be any one of them. Applying this perspective to traditional PD games leads to choices which are cooperative and optimal. Looking at the problem from an impartial point of view leads one to choose the ethically fair and best outcome for the group. It ties the individual's welfare to the group's welfare in such a tight way that the individual has an incentive to choose the strategy which is best for the group. Playing the game from this perspective is the equivalent to the use of an incentive compatible device, hereafter referred to as an "ICD" (Frohlich and Oppenheimer, 1996). The concept of an ICD has been developed by economists to harnessing rational, self-interested behavior to achieve optimal outcomes by finding an institutional structure (such as a tax scheme) that aligns individual interests and group interests. In that way, each individual's incentives correspond to what is needed to achieve the best group outcome.

We ran (1995) a set of 5-person PD game experiments to see whether individuals could learn to solve PD's from the experience of playing from an impartial point of view. We found that the effect of an ICD was negative for solving regular PD's. It appears that the incentive compatible device, by aligning the individual's self-interest with the group's interest, eliminated the subject's need to address ethical concerns. As a result, they seem not to have been evoked in the subjects' minds and were also less active in later plays of the PD with no ICD. Not only were the decisions changed by the change in the game, the values that were evoked were also sensitive to the way in which the choice system was framed. Ethical concerns and motivation were redundant and muted in the presence of an incentive compatible device.

So the results of these experiments are quite different from those involving questions of distributive justice. In the latter, from a rough analysis of the transcripts of discussions, subjects seemed moved to consider the ethical aspects of various principles of distributive justice and to choose on that basis. Frequently subjects commented to the experimenter that they had learned a great deal from participating in the experiment. Their motivation for choosing seemed a mix of egoism and concern for justice. In the PD experiments the impartial choices approach what was best for the group, but they appeared to be motivated solely out of self-interest. Can we identify the fundamental differences in those two types of experiments so that we can judge whether we are likely, in any particular structuring of a choice situation, to succeed in invoking ethical considerations?

Impartiality *Plus* Meaningful Engagement in Discussions

The experiments involving distributive justice decisions were different from the PD experiments in striking fashion. "Behind the veil of ignorance" when subjects were discussing and struggling with the issues involved in a fair principle for distributing income, they were, in general, (apparently) passionately involved. Some discussions lasted almost two hours, with subjects willing to go on until they reached unanimity. Almost no groups showed impatience. Subjects thrashed out a number of substantively important issues and weighed conflicting ethical principles. Should people be allowed to starve? What happens to incentives to work when people are guaranteed an income? Is it fair to tax hard working high earners to support the indigent? The subjects argued about those questions and assumed a variety of points of view. What if I were rich? What if I were crippled? They assumed a range of perspectives. By contrast the environment in the impartial PD's was impoverished. The choices facing the subjects were relatively straightforward and the imposition of impartial play made the best choice quite obvious. There was, by design, no tension between individual and group interest. As a result discussion (in the treatments in which it was allowed) was

short, and peremptory. “We should all just contribute - it’s the best thing for all of us.” was the substance of almost all discussions which usually lasted less than five minutes.

Conclusions

The theoretical lessons we can take from these experimental results are that one *can* construct contexts for evoking ethical values. One way is to design constructs which use impartiality. But impartiality is not sufficient. To do so successfully, one must be sensitive to the context of the decision environment. Impartiality alone will not suffice to evoke ethical values and identify what may be the optimal outcome. The individuals must be engaged in meaningful discussions to draw out the “appropriate” value set. Indeed, some incentive compatible devices which employ impartial reasoning suppress ethical thinking. At the practical level, this poses problems for large scale representative democracies. Can one find ‘field usable’ mechanisms which will encourage the emergence of values and which can generate better societal outcomes? Beginning to understand how individuals can be induced to reason impartially in a fashion which elicits their broader concerns for others may help us achieve better social outcomes.

References

- Audi, Robert (1997). *Moral Knowledge and Ethical Character*. Oxford University Press: Oxford, NY.
- Damasio, Antonio R. (1994). Descartes' Error : Emotion, Reason, and the Human Brain New York : G.P. Putnam.
- Feldman, Fred (1978). Chapter 6 “Egoism.” in *Introductory Ethics*. Prentice Hall: Englewood Cliffs, N.J.: 80-96.
- Frohlich, Norman and J. Oppenheimer, w Pat Bond and Irvin Boschman. (1984) “Beyond Economic Man.” *Journal of Conflict Resolution* v. 28, no. 1, March, 1984: 3-24.
- Frohlich, Norman, Joe Oppenheimer, and Bernard Moore (2001) “Some Doubts About Measuring Self-Interest Using Dictator Experiments: The Costs of Anonymity,” Journal of Economic Behavior and Organization Vol. 46, No. 3(November): 271-290.
- Frohlich, Norman (1992), “An Impartial Reasoning Solution to the Prisoner’s Dilemma.” *PUBLIC CHOICE*, 74. No. 4 (December): 447-460.
- Frohlich, Norman [and Joe A. Oppenheimer]. (1995) The Incompatibility of Incentive Compatible Devices and Ethical Behavior: Some Experimental Results and Insights. Public Choice Studies, V. 25: 24-51. (Incorrectly published without Oppenheimer’s name on it)
- Frohlich, Norman and Joe A. Oppenheimer (1996).”Experiencing Impartiality to Invoke Fairness in the n-PD: Some Experimental Results.” *Public Choice*, 86 (117 - 135).
- Gert, Bernard (1998) Morality: Its Nature and Justification. Oxford University Press. Oxford, England.

- Grether, David M. and Charles R. Plott, (1979) "Economic Theory of Choice and the Preference Reversal Phenomenon," American Economic Review, 69 (September): 623 - 638.
- Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. 1994. Preferences, Property Rights and Anonymity in Bargaining Games. Games and Economic Behavior. 7(3): 346-380.
- Hoffman, Elizabeth, Kevin McCabe and Vernon L. Smith (1996). "Social Distance and Other-Regarding Behavior in Dictator Games." American Economic Review. V. 86, No. 3 (June): 653 - 660.
- Jackson, Michael and Peter Hill, (1995) "A Fair Share," Journal of Theoretical Politics 7 (2, April) 69-179.
- Kahneman D. and A. Tversky (1982), "The Psychology of Preference," Scientific American, 246 (January): 160-173.
- Lissowski, Grzegorz, Tadeusz Tyszka and Włodzimierz Okrasa. Principles of Distributive Justice: Experiments in Poland and America. Journal of Conflict Resolution v. 35, No. 1, March, 1991: 98 - 119.
- Nagel, Thomas (1991) Equality and Partiality. Oxford University Press. Oxford, England.
- Nozick, Robert (1974) Anarchy, State and Utopia, New York: Basic Books.
- Oleson, Paul E. (1997) "An Experimental Examination of Alternative Theories of Distributive Justice and Economic Fairness." Paper presented at Public Choice, San Francisco, March.
- Roth, Alvin, E. (1995) "Bargaining Experiments," in Kagel, John H. and Alvin E. Roth, eds. (1995) THE HANDBOOK OF EXPERIMENTAL ECONOMICS. Princeton, Princeton University Press: 253-342.
- Quattrone, George A. and Amos Tversky, (1988) "Contrasting Rational and Psychological Analyses of Political Choice." AMERICAN POLITICAL SCIENCE REVIEW. (82, NO. 3 SEPT.) 719-736.
- Saijo, Tatsuyoshi, Shusuke Takahashi, and Stephen Turnbull (1996). Justice in Income Distribution: An Experimental Approach. Mimeo: presented at 1996 ISA, San Diego, April 18, 1996.
- Tversky, Amos, S. Sattath and Paul Slovic. (1988) "Contingent Weightings in Judgement and Choice." Psychological Review 95, 3, 371-384.