Comments & suggestions are welcome
For resubmission to the J of Theoretical Politics
Monday, May 23, 2005

Skating on Thin Ice: Cracks in the Public Choice Foundation:[1]

Norman Frohlich *(Corresponding author)*
I.H. Asper School of Business
University of Manitoba
Winnipeg, Manitoba R3T 5V4
(204) 474 6385; e-mail: frohlic@ms.umanitoba.ca

&

Joe Oppenheimer
Department of Government & Politics
University of Maryland, College Park, Maryland 20742
(301) 405 - 4113; e-mail: joppenheimer@gvpt.umd.edu

Key Words: rationality preference self-interest cognitive model

Abstract:

The behavioral assumptions upon which Public Choice, and Game Theory are built, are false. This may not seem to be 'too serious' since many theories are only approximations. But a broad spectrum of experimental results has shown the problem to be greater than one of workable proportions and discuss its implications for knowledge claims and theory construction. We illustrate the problems by examining some anomalous results in both voluntary contribution games and dictator games. We argue that there are difficulties with both the content and structure of the preferences posited in the standard model, sketch the outlines of a possible solution, and discuss some of the implications of this new perspective.

1. Thinking about this problem began very early in our joint careers, and was pushed further by data we collected in the late 80's and mid-90's, largely funded by the Social Sciences and Humanities Research Council . It has been given new voice with the help of an NSF grant, and the collaborative arguing with Elinor Ostrom and TK Ahn. Our arguments have been presented at meetings of the Public Choice Society and at an NSF funded conference held in Indiana in January, 2003. This paper is a natural outgrowth of those conversations.

Skating on Thin Ice: Cracks in the Public Choice Foundation
Outline / Contents

Tables and Figures

Skating on Thin Ice: Cracks in the Public Choice Foundation

Norman Frohlich and Joe Oppenheimer

The field of Public Choice has been built on two behavioral assumptions of micro-economics: rationality and self-interest. The underlying article of faith has been that these assumptions allow one to explain political behavior. In political science, this endeavor has been dubbed "rational choice theory." That theory has cast a good deal of light on a number of political phenomena. Models have been developed (often using constructs from game theory) to explain a variety of non-market interactions and those models have been useful in developing effective policy alternatives for handling social problems.

But there has been a fly in the ointment: some stubborn facts. Over the past few decades this has become ever clearer. The behavioral foundations of micro-economics have suffered a few direct hits. First, psychologists chipped away at the simple rationality postulates with experimental evidence of preference reversals, and other anomalies, often lumped together as 'prospect theory.' Second, experimentalists from a variety of disciplines in the social sciences, have identified limits on the assumption of self-interest. (Berg, Dickhaut and McCabe, 1995; Bolton, 1991; Dawes and Thaler, 1988; Dawes, van de Kragt, and Orbell, 1990; Eckel, and Grossman, 1996; Fehr, and Schmidt, 1999; Frank and Schulze, 2000; Frohlich, and Oppenheimer, 1984; Frohlich, Oppenheimer, and Kurki, 2004; Hoffman, McCabe, and Smith, 1996; Kahneman, Knetsch, and Thaler, 1986; Larrick, and Blount, 1997a; Saijo, and Yamaguchi, 1992; Sweeney, 1973) Given these problems, the chains of reasoning from micro understanding of decision making at the individual level, to macro group and institutional outcomes has been called into question. So while policy applications have been forthcoming and even adopted at the macro level, the logical structure of the evidence justifying those applications has become increasingly problematic. The anomalies call for a reconsideration of the micro foundations of this purportedly successful enterprise and for strengthening the links in the theoretical arguments.

In this paper, we give a few examples of the problems that have emerged. The problems illustrate the palpable lack of fit between the data and the model's predictions and call out for revisions that can yield better predictions. We characterize the problems as flowing from an imperfect understanding of the empirical content and structure of preferences. Rationality and self-interest, as we have come to know and love them, may not be what they seem. We raise concerns about the separability,[1] stability, and uniqueness of preferences, particularly when explaining other-regarding behavior. We sketch an alternative perspective which includes multi-faceted, non-separable preferences that are context dependent. The strengths and weaknesses of this approach are discussed, and we nest the discussion in the epistemic status of modeling as an instrumentalist, or a realist exercise.

Introduction

The standard theory can be depicted quite simply. At the base of the pyramid are two heavy duty working premises: Individuals are *rational* and *self-interested*. Rationality is usually defined as choosing the best alternative: maximizing. This requires unique, stable, well ordered preference

1. Separability of preferences refers to preferences being non-interactive among individuals. Some of the implications are spelled out in chapter 2 of Sen (1973).

structures.[2] So entrenched was this idea that when we went to graduate school, and through the 1970's, most sophisticated political scientists and many economists, presumed rationality to be a tautology: how could it be otherwise? Of course, much earlier, Kenneth May had shown that the premise was *not* a tautology (May, 1954). Individuals could be caught with inconsistent, intransitive, and cyclic preferences, thereby creating confusion. Somewhat in anticipation of, and in other cases in response to, such empirical findings, Simon (1947, 1992), and Lindblom (1979), pushed for the notion of what was called 'bounded rationality.' If individuals did *not* maximize (or at least not globally) the premises could be reformulated more realistically to preserve local maximization, or bounded rationality. Zeckhauser and Shaefer (1968) showed the logical equivalence of bounded rationality with maximizing behavior when information is costly. That seemed to put the problem to rest. But down the hall, in departments of psychology, others were finding bigger problems. Individuals didn't display the consistency required by the assumptions even when they had full information. *Consistent* maximizing might not be a part of human nature after all (see the good summaries of the findings in Rabin, 1998; Quattrone and Tversky, 1988; Grether and Plott, 1979; Simon, 1986; Tversky and Kahneman, 1986; as well as Shafir and Tversky, 1994).

The second premise of most rational choice models, *self-interest*, had also been argued to be tautological. But those arguments were also based either on stipulative definitions that didn't coincide with rational choice usage, or have been found to be wrong. Empirically, in the market contexts to which the 'standard theory' had originally been applied, self-interest appeared to be relatively robust. But as rational choice theory was applied to areas ever more distant from markets, the assumption of self-interest proved more problematic; it did not predict the observed results. (See Frohlich and Oppenheimer, 1984, 2000, and a variety of dictator experiments: the literature is reviewed in Roth, 1995).

Perhaps the area most fraught with anomalous findings grew out of one of Public Choice's earliest and most celebrated triumphs. In 1965 Mancur Olson boldly analyzed the collective action problem in rational choice terms. He succeeded in explaining why groups of rational self-interested individuals often failed to act to achieve an attainable group benefit. As he put it in the third paragraph of the Introduction:

> But it is not in fact true that the idea that groups will act in their self-interest follows logically from the premise of rational and self-interested behavior. It does *not* follow, because all of the individuals in a group would gain if they achieved their group objective, that they would act to achieve that objective, even if they were all rational and self-interested. . . . The notion that groups of individuals will act to achieve their common or group interests, far from being a logical implication of the assumption that the individuals in a group will rationally further their individual interests, is in fact inconsistent with that assumption.

Sometime later, empirical tests of Olson's insights, first in the field (see for example the discussion in Baumol and Oates, 1979), and subsequently in laboratories (Ledyard, 1995, is a fine starting point) failed to bear out the drastic, and pessimistic corner result of non-contribution to group goals. Rather what appeared was a consistent pattern of partial contribution in what became to be known as 'voluntary contribution mechanisms' or VCM's. Attempts to explain consistently higher levels of individual contributions than the zero levels predicted by game theoretic models first

---

2. Sen (1970) along with others, shows clearly why the presumptions of well ordered, preferences and maximizing behavior require some assumption like transitivity. Less demanding definitions are developed there, with the trade offs specified.

clutched at subject error, or ignorance (see Saijo and Yamaguchi, 1992 for a particularly clever attempt to explain the anomalous behavior). Andreoni, (1995) and later Houser and Kurzban, (2002) have also shown that confusion can play a role in explaining anomalous behavior, but the explanations have been only partially successful. Slowly, but surely, the experimentalists have begun to focus more directly on the problem of the foundations of the theory. To sharpen the focus on the problems with the standard model we start with an examination of some data that is characteristic of the anomalies found in VCM's.

Some Preliminary Evidence

Micro-Data and Collective Action

A variety of repeated-play VCM experiments have yielded a consistent pattern of results. One result could be characterized as significantly higher group contributions than the theory predicts: groups generally begin with contributions well above the predicted level of zero, but that declines over time (roughly from 40% or 50% to 25% or 20%, but only rarely approaching the corner solution of zero contributions). So there is a problem in predictions at the macro level of the group.

Some researchers, looking at macro patterns and summaries of group behavior over time, have generously interpreted this consistency as giving one hope that the theory is partially supported by the data. Often the analysis stops at that point: research question asked, test designed, completed and reported. The results are implicitly assumed to be "close enough." Had they, instead, interpreted the results as "not close enough" they would have asked the question: "What's going wrong?"[3] And to answer that question they could have gone deeper into the micro data available to them: the data about the choices of individual subjects in their experiments. Of course, at the individual level there is much more noise in the data than at that group level. How that noise reflects complex individual behavior and how that behavior might require reformulation of the foundational premises discussed above are the subjects of this paper.

To illustrate the depth of the problem, consider some data from VCM experiments that we have run (Frohlich and Oppenheimer, 1996 and Frohlich and Oppenheimer, 2003b).[4] The experiment is quite typical in its parameters (see Table 1): the size of the group is 5, each individual is endowed with a budget of 10, and must decide the proportion of the budget to be allocated to a shared or public good that yields a return of 40% to each individual. The decisions are

Table 1: 5-Person Prisoner's Dilemma (Showing Payoffs Only to One Player)

	Amount Given by Others				
1 Person's Strategies	**40**	**30**	**20**	**10**	**0**
give 0	26	22	18	14	10
give 10	20	16	12	8	4

3. Some few have, at least at times, raised that question. They certainly include a number of experimentalists including James Andreoni, Rachel Croson, Katherine Eckle, Charles Plott; T.K. Ahn; Elinor Ostrom, Saijo, Mark Issac, James Walker, and Arlington Williams; along with others.

4. Protocols for these experiments are available upon request.

repeated (with the same group) 7 or 15 times (depending upon the particular experiment).[5] The standard theoretical prediction from the theory of collective action, is that individuals will choose their dominant strategies: giving nothing (top row in the table). Later work in non-cooperative game theory indicates that other equilibria can be sustained (the folk theorem); but without communication the non-cooperative Nash outcome of all playing their dominant strategy is most compelling. And as noted above, this contradicts the standard experimental result: a non-corner solution of higher levels of contributions, decaying over time, but not to zero. And virtually no subjects consistently donate nothing.

The graph in Figure 1 display some typical results for five groups (aggregated)over the 15 rounds of repetitions. The data does not bear out the strict theoretical prediction of zero contribution, but does conform to the results usually found in experiments of this type.

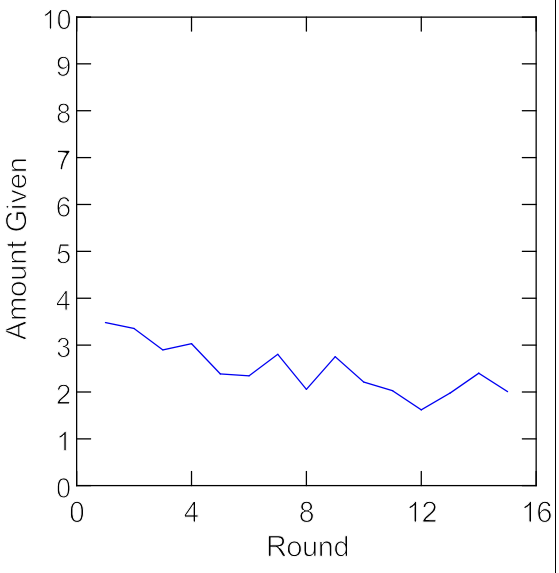

Figure 1: Average contributions of 5 person groups in a 'typical' VCM experiment

So how can we explain what is happening? The assumptions of rationality and self-interest are certainly not giving us leverage at the micro level of individual decisions, nor are they supportive of the observed group level outcomes. How does this result fit with either the idealist notion that we are reasoning animals who support our collective needs or the "realist" theory of selfish rational choice that argues quite the opposite? What sorts of *individuals* are behind these results? Are there differing types of individuals: some give and some don't? Is it the case that, for example, 60% say, "Not me! I'll be no one's "sucker" and 30% say, "Yes, count me in!" And how do people alter their behavior? Do a few individuals (say 10%) slowly move towards non-contribution? Perhaps there are just these three types: the rational egoists - equilibrium playing self-interested responses, the unconditional cooperators (both of these sorts would have a response line that would be horizontal or constant over the course of the experiment), and the learners. The decline would come from learning: a switching of strategies over time among the few learners.

But other behavioral patterns could also generate these results. There may be conditional cooperators and reciprocators as conjectured by Rabin (1993); Cain (1998) and Croson (1998) - individuals who are learning, or trying to get something going by signaling, or trying to contribute when it could make a difference (and their trajectory of behavior would look quite different in this graphical space; see Janssen and Ahn (2003) for an interesting model of signaling behavior in this environment). And some may be reacting to others' non-cooperativeness: attempting punishment by withholding contributions, even though in this simple VCM design there is no way of directly targeting punishment to a particular individual - see (Ostrom et. al. 1991 and 1992). We could divide the conjectural literature regarding the anomalous results of VCM's into 3 groups:

5. The subjects don't know when the experiment will end.

- Some consider the changes as coming from subjects' learning about the choice environment but introduce no explicit heterogeneity of behavioral types;[6]

- Others consider the possibility that we can identify types of individuals;[7]

- Still others, see Chaudhuri, Maitra and Graziano, 2004, consider variation of behavior both between types of individuals and via conditional behavior within individuals over time.

Considering individual types leads us to shift our focus from the groups to the individual. How many simple egoists, or unconditional cooperators are there in a typical data set? What is the proportion of non-conditional actors: actors whose behavior can be depicted by horizontal lines over the time of the experiment? We have not checked this in any large sample of data, but looking at the 25 individuals who participated in the VCM experiments with 15 rounds, what do we find?

Table 2 presents the data, and line one shows that of the 25 subjects, *only two* are unconditional non-cooperators (rational self-interested egoists) who never contribute anything. There are no unconditional cooperators. That leads us to the next step: can we inductively arrive at a 'typology" of actors who have interpretable behavioral patterns? While no other stark demarcation lines appear in our data, we can roughly identify and count some possible types of conditional behavior. We might characterize token responders as those whose median giving is below 20% of their endowment;[8] moderate responders are those who give 50% at least twice but never give their full endowment; high responders are those who give 100% of their endowments at least once.

Each of these types can be exemplified graphically (see Figures 2 - 4 where we show examples of these types). Obviously, the distinctions and especially the cut-offs between them, are arbitrary: we have nothing but inductive conjectures for all but the pure egoist 'type.'

Table 2: Distribution of Responses for 15 Round Observations

types	Percent	total
egoists	8%	2
token	32%	8
moderate	36%	9
high	24%	6
Totals	100%	25

Nothing about the analysis or graphical representation is meant to reify these categories but they give a possible 'cut' at the 'type' problem we face: how to reconcile divergent (and often quite

6. For an example, see Issac, Walker and Willams, 1994 who play with various conjectures to close the gap between theory and observation. For example, they test the conjecture that poor information can explain the differences (it doesn't); that allowing for simple learning over time could lead to better predictions (it doesn't); and finally, they consider such preference altering possibilities as a 'warm glow effect.'

7. See for example, Croson, 1998; Gunnthorsdottir, et. al., 2001; Ashley, Ball and Eckle, 2003; and Janssen and Ahn, 2003, all of whom try to develop an inductively justifiable set of types of individuals. Croson, for example, argues that virtually 80% of individuals behave, in some fashion, reciprocally.

8. One "token" responder gave the full endowment on the first round but then only contributed once (and then only 20%) in the 14 rounds thereafter.

erratic) behavior across types of individuals with the findings of relative uniformity at the macro-level of outcomes.

Looking at the data, we note that 1) there are no other simple non-conditional players (that always give a set amount) other than those that give 0; 2) typical behavior of others (see Figures 2 - 4) is quite saw-toothed: behavior not easily interpretable as a product of simple 'learning' alone. So, what is the reality behind behavioral types? Many conjectures are possible. Inconsistency can stem from erratic learning, complex interactive cueing of conditional behavior among the subjects in the experiments, general instability of preferences, a lack of care in the decision making, or violations of self-interest.

Considering a Few Possible Explanations

Factors Affecting Individual Decisions in a VCM

Let us consider this last possibility in a bit more detail. Self-interest is well defined; but what of its denial? There are a variety of modes of non-self-interested behavior (Frohlich and Oppenheimer, 1984). Generalized models of rationality, incorporating non-self-interested behavior in plausible, flexible ways have been developed (see, for example, Fehr and Schmidt, 1999). Empirical investigations of forms of non-self-interest have recently gained steam in the laboratory. Many carefully constructed experiments have investigated non-self-interested behavior (see Berg, Dickhaut and McCabe, 1995; Bolton, and Zwick 1995; Dickhaut, Hubbard and McCabe 1995; Eckel, and Grossman, 1996 and 1996; Frohlich, Oppenheimer, and Kurki, 2004; Hoffman, McCabe and Smith, 1996; Hoffman, McCabe, and Smith, 1996; Hoffman, McCabe, Shachat, and Smith, 1994; Kahneman, Knetsch, and Thaler, 1986; Konow, 2000; and Ruffle, 1998; Andreoni, 1995). The growing body of evidence supports the common sense observation: human behavior is far too complex to be characterized, in these contexts by the extreme construct of self-interest. This has led many of us to try to explain behavior with more general assumptions defining more complex utility functions (see Cain, 1998; Cox, 2001; Fehr and Schmidt, 1999; Frohlich, Oppenheimer, and Kurki,

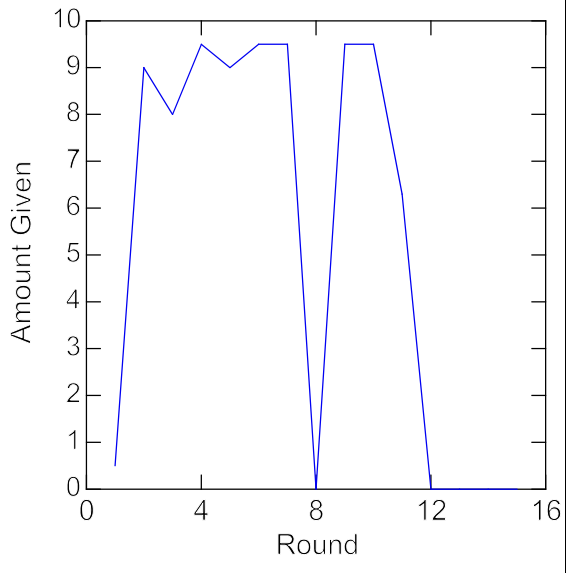

Figure 2: Example of a High Contributor to the Bonus Fund

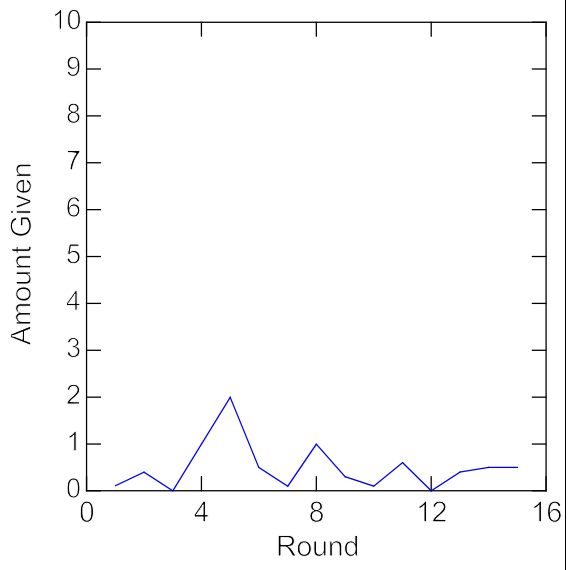

Figure 3: Example of a 'token' contributor to the Bonus Fund

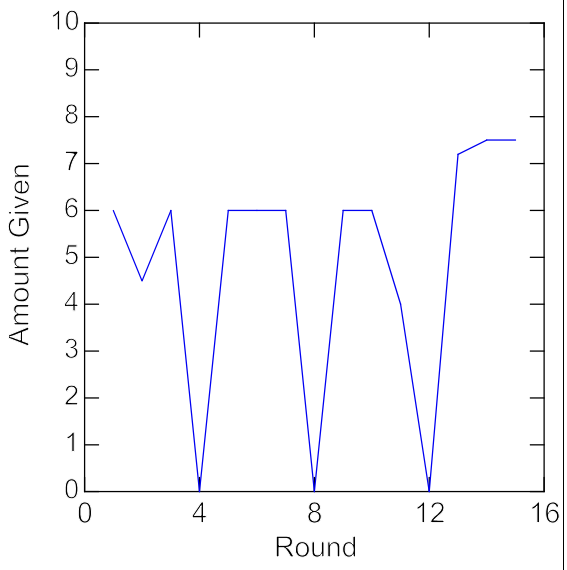

Figure 4: Example of a 'moderate' contributor to the Bonus Fund

2004; Rabin, 1993.)

For example, some non-self-interested preferences can be represented by a utility function that is positively responsive to others' welfare or preferences. The precise identity of these interaction terms and their functional form can furnish a partial expression of the patterns of behavior exhibited by our rough "types." It would explain why some individuals (with strong terms of this sort) display more "ethical" motives or express more concern about the welfare of others and contribute more than those who lack terms of this sort, or who have lower weights attached to those terms.

However, any such concern, by itself, cannot account for the spikey up and down fluctuation in contributions shown by many of the individuals depicted (and in the other individuals not depicted). Something additional is required for that.

Table 3: Ethical Concern by Type of Contributor

Type of Contributor	Mean "Ethics" Score	Standard Deviation	N
Egoist	8.50	12.02	**2**
Token	65.50	64.82	7
Moderate	123.56	31.38	9
High	156.60	29.00	5

Obviously what is called for is a type of behavior conditional upon the individual's understanding of the behavior of others (Rabin, 1993; Cain, 1998). Two sorts of responses could be contrasted to explain this apparently erratic behavior (and the reader may, no doubt, supply others): First, individuals may be trying to do as well as they can: "signaling" their *bona fides*: contributing to encourage future cooperation. This sort of behavior, either self-interested, or not, could well become cyclic or saw-toothed, inasmuch as communication via the "signal" of contributing varying amounts will be highly imperfect. But individuals may be concerned about their payoffs, in other ways, for example, their payoff *relative* to those of others in the group. So, for example, if *i* were to contribute a reasonable amount in some round, while others contributed little, *i*

might feel hard done by and reduce her contribution in the next round. Desires not to be taken advantage of may enter into the calculation and make individuals reactive to others' behavior in previous rounds. Those with such non-separable desires may actually want to punish those who took advantage of them by withholding contributions. On the other hand, they may be more than willing to donate when they feel others contribute.

We can test for the positive concern for others using two of the questions from a post-experiment questionnaire tapping into subjects' ethical concerns. In an 'exit' survey, we asked subjects how concerned they were about both acting fairly: "Doing my fair share was important to me." and about their concern about the payoffs to others: "Concern about the payoffs of others was important to me." Their responses, scaled on a continuous line, were aggregated into an index that we entitled "*Ethics*." The four types of subjects described above showed marked differences in their scores on this ethics index (see Table 3). An analysis of variance of "*Ethics*" as a function of type explains 74 percent of the variance and is significant at the .001 level with an F value of 17.59.[9]

Table 4: Illustrative Regression Explaining Contributions

n = 972 R = 0.646, R^2 = 0.417, Adjusted R^2 = 0.416

Effect	B	S E	$	t	P(2 Tail)
Constant	-1.35	0.212	0	-6.36	0.001
Ethics	0.018	0.002	0.305	11.99	0.001
Others' (lagged) giving	0.157	0.008	0.493	19.34	0.001

But we can also look at the effects of conditionalized behavior such as signaling and negative reactions to others' low contribution rates by tracking contribution levels as a function of others' most recent contributions. Signaling, being willing to contribute conditional on the behavior of others, and a desire to punish others, (or not be taken as a sucker), would all relate current choice to prior experience: most plausibly, to what others collectively have done in the immediately previous round. We can see how well the combination of positive concern for others' payoffs, plus the conditionalized behavior towards others' contributions can explain the swings in round by round levels of giving. To get traction at that level of detail we switch to the 7 round cases (where we have many more observations N= 175 persons, making more than 1000 decisions). Such a simple two-variable model (see Table 4) using the ethical orientations of the experimental subjects (labeled "Ethics") along with a measure of the total monies contributed in the *previous round* by everyone else (labeled "Others' (lagged) giving") explain more than 40% of the variance, but this still leaves considerable variance unexplained.[10]

9. Gunnthorsdottir (2001) reports that free-riders have dispositions that are differentiated from cooperators. We are reporting something quite akin to that.

10. One outlier is removed from this analysis. Details of the questionnaire, etc. are available in the original study (Frohlich and Oppenheimer, 1996).

Naturally, there are other ways to look at individual responses. Table 5 shows the possibly powerful effect of a single egoist in the group. When *one* of the players in the group is a uniform defector, the extremely pessimistic result of no contribution, predicted by the Olsonian model, is borne out.

Table 5: An Egoist Spoils the Party

types	Groups 1	2	3	4	5	total
egoists			1	1		2
token		3	2	2	1	8
moderate	2	1	2	1	3	9
hi reasonable	3	1		1	1	6
grp outcomes in the 15 round experiments	fluctuates around 5 = rights situation after collapse	bad experience in round 8, then ok slowly off	floats to 0	floats to 0	fluctuates around 2.5	25

Evolutionary game theorists can rejoice in this as a confirmation of a generalized argument by Levati (2003). She argues that evolutionary responses by conditionally cooperative subjects to a pure egoist in a VCM group would often result in the complete collapse of cooperation. No-one really conjectures that evolutionary time speeds by so quickly that in the brief 7 or 15 rounds of an experiment that an evolutionarily predicted result would magically appear. On the other hand, we may have evolved a visceral response to blatant free-riding, which is evoked by such egoists, and prompts us to defect. (See the work by Camerer and Quartz discussed in Adler, 2004). Nevertheless, the existence of a single egoist could provide another complicating factor in those cases.[11] But in groups without an egoist, we still are far from explaining the patterns of individual behavior. And, to repeat, a parametric level of ethical concern can not explain why there are such variations in giving over rounds.

Returning to the substantial variance left unexplained at the individual level by the regression (see Table 4), we must conclude that something else is causing behavior that remains unexplained. The relatively high r^2 (by social science standards) does not fully explain what appears to be a typically saw-toothed pattern of individual behavior. And a shift of levels of analysis to the single group (not the aggregation of all groups) reflects a similar but somewhat dampened pattern of saw-toothed behavior at the group level, as compared to the individual level. Figure 5 gives an example (in this case of a rather successful group). This pattern can also be contrasted with the aggregate behavior of all five groups represented in Figure 1 with its smaller swings.

11. In a supportive finding, her experiments, Gunnthorsdottir (2001) notes that cooperators who interact more with free-riders than with other cooperators will contribute less to public goods.

The successful regression results, with ethical concerns and a lagged response variable, hide the fact that there are still great unexplained oscillations in behavior and outcomes. So we are left with the question with which we began: how do we explain collective action on the basis of individual behavior?

One option would be to introduce learning. But any simple learning model would be handicapped because the environment of each of the subjects is continually shifting as each individual tries to learn a lesson from the behavior of all the others. And the direction of learning would be a function of at what level and when they started contributing.

So the simple model of the strategic interaction of the game, even with some consideration of 'types' of individuals, and versions of non-separable preferences and conditionalized behavior is not capturing the complications of the behavioral choices involved.

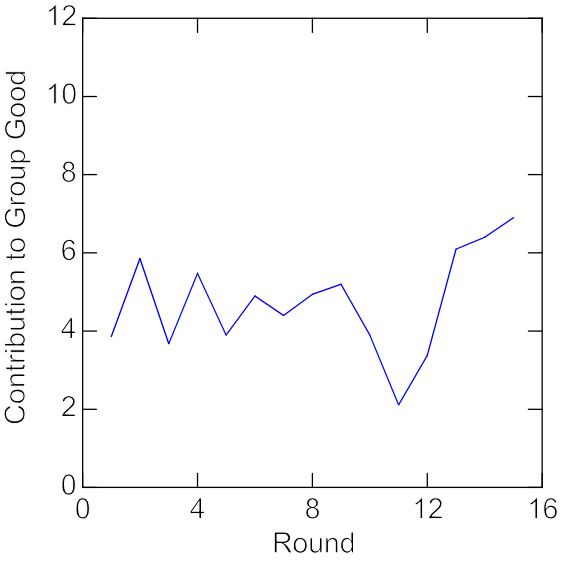

Figure 5: Saw Toothed Group Behavior

A fundamental problem of the group situation is that no player is in a position to interpret others' actions unequivocally. For example, is someone's zero contribution selfish or punishing? And by the same token, how can a player understand the effects of their behavior on others? Moreover, *all* players are making decisions at the same time, and contrary actions by players: one, contributing more, perhaps to signal cooperation, and another withholding any contribution to punish or to signal displeasure, can cancel one eachother out. So, clearly there is a great deal of noise in the system. That may help to explain why so much of the variance remains unexplained by the variables "ethics" and the "prior contributions of others" in the regression reported above. In sum, the seemingly erratic behavioral results appear to stem from the subjects' reactions to the perceived behavior and intentions of others: developing a somewhat unpredictable pattern.

There is body of evidence that may support the notion that players are individually seeking the fully cooperative Pareto optimal core of the game but "signaling and punishing" are impeded by the noise in the system. One of the most robust findings in the VCM literature is that face to face (and to a somewhat lesser extent, electronic) communication is extremely effective in inducing cooperation. Such communication allows for "signaling" that is independent of the choice act (and as such has often theoretically dismissed as 'cheap talk'). In other words, the erratic behavior can be eliminated if subjects are allowed to communicate (Isaac and Walker, 1988; Frohlich and Oppenheimer, 1998; Ledyard, 1995, Ostrom, 1998). This gives force to the induction that the variance in behavior at least partially reflects an attempt to communicate. Just to illustrate, in an alternative treatment, given the possibility of face to face communication in the first 7 rounds of the VCM experiments discussed above, behavior was very different. It led to almost universal contributions of the full $10 by everyone in every round (mean contribution level 9.99, S.D. 0.047, n=30).

This leaves a final question: "Why do contribution levels not decay to zero?" The answer, according to many, lies in the reward structure and in a sort of satisficing response to it. At about

20% contributions, in our game, a five-person group has $10 contributed, that generates a group benefit of $4 per person. Such an outcome yields an average net (after contribution) per person payoff of $12: better than the $10 initial endowment which can be preserved by mutual defection. As cooperation decays over time, players may become aware of the fact that continuing to respond in a more selfish fashion may cause the situation to collapse below the point where people are getting (on average) a positive return over the rounds. This would push everyone down to the $10 equilibrium. By continuing to contribute relatively small proportions of their endowment (and perhaps signaling), they keep themselves in position with their heads above water i.e. getting an adequate return. This may be thought of as a kind of satisficing (see Schelling, 1973 who refers to this as a break-even point).[12]

Of course, in the laboratory, a group's ability to keep above that minimal floor may be a function of the distribution of "types of persons" in that group. Perhaps one could roughly predict the types from their dispositional, or "ethical" orientations. Presumably, those with high "ethics" score get non-monetary rewards from contributing, contribute more, and are much more likely to sustain higher contribution levels in the group. Thus, the final round contribution mean need not be the relevant statistic: no one is necessarily motivated by or responding directly to them. To reify it by explaining it as the outcome of group behavior is bizarre. Rather, some groups are doing better, others worse, than 20%: some receive nothing, and perhaps should be analyzed separately.

Revising the Behavioral Assumptions

This discussion of the factors affecting individual decisions in a VCM call for a revision of the underlying behavioral assumptions so as to reflect non-separable preferences and conditional responses. One starting point could be the Fehr/Schmidt model (1999). The Fehr-Schmidt model characterizes an individual's valuation of a payoff of x_i to one's-self as compared to a payoff of x_j to another as the payoff minus the valuation (cost) of any inequality:

$$U(x_i) = x_i - \text{"} \, \text{Max}(x_j - x_i, 0) - \$ \, \text{Max}(x_i - x_j, 0) \qquad \textbf{(1)}$$

Fehr-Schmidt refer to the values behind this model as inequity aversion: **"** indicates the weight attached to the loss one suffers when the other party gets more than oneself, while **$** stands for one's loss from inequality (assumed to be disliked) when one gets more than the other. While this utility function relaxes the self-interest assumption, it does so in a narrowly particular way. It assumes all differences are valued negatively (both **"** and **$** are assumed $0, and **"** >**$**). It may be that some individuals are not troubled (they may even be pleased) when they receive more than some others do (i.e. **$** # 0). This would be true if the other in question were deemed worthy of punishment. And certainly the value of both **"** and **$** may be conditional upon the prior intentional action of one's self and the other individual. So, for example, if *i* feels that he has done 'the right thing,' and the other (*j*) has not behaved acceptably by intentionally not contributing "enough" over

12. A number of clues exist why this isn't a problem in the world outside of the laboratory. Chaudhuri and Graziano (2004), for example, show that with 'intergenerational' advice, there is a strong tendency for groups to do better. Without such cues, in the laboratory, we are facing a more difficult explanation. Croson's discovery of high levels of reciprocity can be combined with two other recent discoveries. Ashley, et. al. (2004) have found that many reciprocating individuals quickly reduce their contributions if they find they have contributed more than average in the previous round. (For those who contribute less than average, the response is not significant.) Such a response could lead to behavior asymptotic to a variable, significant, non zero value of contributions. The Isaac et. al. (1994) solution was to examine the Schelling notion a bit more formally. If subjects were oriented toward 'doing at least as good,' as some low break even point, one can develop expectations that can sustain the threshhold.

several rounds, the utility impact on *i* of getting more than the "cheater", *j* , in the next round, may actually be positive rather than negative and be reflected via a negative **$**, and an inflated **"**. If **$** is more negative than -1, it would imply a willingness to bear a cost to punish a cheater. On the other hand, in a larger group (say 3 persons, in a VCM) if that same *j* were to have been a consistent contributor, who happened, in the course of events, to have received a higher payoff (of the same magnitude) as a result of trying to punish some other "cheater" *k*, *i* might well not begrudge *j* that "justifiable" differential. The effective size of **$**, might be 0: assessing *j*'s extraordinary gain as a cost of sending *k* a message and teaching him a lesson. In other words, the values of the parameters within a given player's utility function may vary across rounds conditionally not only on the other's previous contribution levels, but also on the *intentions* of that player which motivated the behavior. This analysis implies that differences in payoffs matter, but that the reasons for those differences also matter (Rabin, 1993 and Cain, 1998). This seems to indicate that there may be normative aspects of behavior that impact players' preferences that are not captured by the simple Fehr/Schmidt model. Yet such normative components are likely to be needed for modeling behavior in the relatively simple VCM experiments. A more complex version of their model, (with a more complex conditional structure) or a different model altogether seems to be called for.

The role of context on conditionalized non-separable preferences

Modeling preferences that are conditionalized, not just on the actions of others, but also on their intentions, poses complex informational problems. After all, intentions are not observable. Ultimately, each player must make up her mind about the intentions of others and act in accordance. In some situations sanctioning may be called for, while in other situations acceptance of a similar outcome with differing imputed intentions, may be acceptable. The choice turns on whether the outcome is a result of a normatively acceptable behavior. To complicate matters, in some contexts, some norms are important (i.e. expected to be adhered to) while in others, the same norms are viewed as irrelevant. A measure of cooperation may be expected among five players in a VCM and deemed worthy of punishment if not adhered to. However, other mathematically similar situations, might be governed by differing norms. For example, consider five bidders, failing to cooperate in an auction to keep the price (a public good to that group) low is obviously similar to a VCM, but it has different governing norms. In other words, context, and intentions matter and interact. Or, to put it more strongly: context may determine how intentions matter. Norms are context dependent.

So, if context determines which elements enter into an individual's decision structure, then explanations of behavior require that we identify choice contexts.[13] Indeed, the experiments that resulted in Prospect Theory (Tversky and Kahneman, 1981), demonstrated that the *stability* of preferences, and hence individual choices, are sensitive to the individual's interpretation of the decision context, and hence, dependent upon the way that the decision problem is framed.

It has by now been widely demonstrated that the mere framing of decisions can lead to preference reversals (or seeming irrationality). We demonstrate an effect of that sort in Frohlich and Oppenheimer, (2003b) for VCM's, but we have also been concerned about the problem in general. To demonstrate how context, can evoke different preference structures, we crafted a set of dictator experiments with a difference (Frohlich, Oppenheinmer and Kurki, 2004). Rather than simply *giving*

13. Although that may appear to be a straightforward matter, a moment's reflection will reveal that it is not. Sometimes contexts are ambiguous, and actors are not sure of the context in which they are choosing. And, of course, a theory of types of 'contexts' will be culturally determined.

dictators money to allocate, dictators and their paired other subject produced income by doing work. Then the dictators allocated the total joint income anonymously. Our conjectures were that the work context would 1) evoke "entitlement" values of a normative nature and 2) lead to more sharing on the part of the dictators than in a normal dictator experiment, and 3) cause the sharing to conform to some normative rules reflecting entitlement.

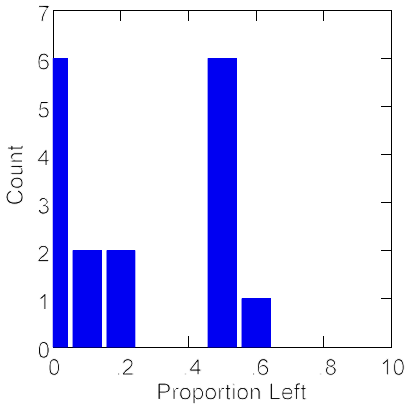

Figure 6: Amounts left in normal dictator experiments

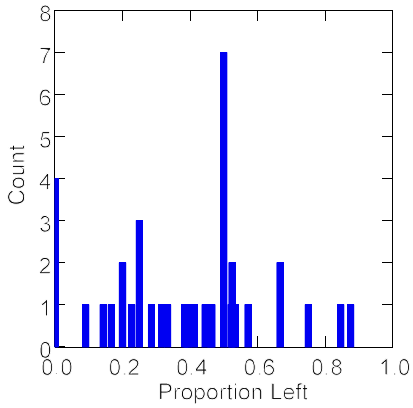

Figure 7: Amounts left in dictator experiments when is product of joint production

Figure 6 shows the distribution of the proportion of $10 left to the individual paired with a dictator in a standard (one room) dictator game (Frohlich, Oppenheimer and Moore, 2001b) and Figure 7 shows the proportion of the joint returns left when the amount to be distributed was the product of both party's work Frohlich, Oppenheimer and Kurki, 2004). It is clear that the distributions are different. The introduction of a new element (production) into the situation has impacted the dictators' choices by (presumably) changing their preference structures. A strict rational self-interested analysis of the two experiments would predict the same outcome: dictators should keep all of the money. But, parallel to the VCM results, that isn't what happens. Different things happen when the contextual element of production is brought into play. In the new production context, new elements are evoked in the dictator's utility function that further modify the self-interest assumption to make the preferences non-separable.

Our experiments permit us to identify the value structures being manifest and test the conjecture that these additional values can be modeled in a fashion using the mathematical form of the Fehr-Schmidt model with changes in the content of some of the terms in their equation. One of the values tapped into in the production context is entitlement. Of course, this is only one model that could be written as a template to explain behavior when there are multiple values individuals might evoke when trying to do 'the right thing' within that context. Without providing the mathematical formalism, the utility function contains the following elements:[14] Each person was assumed to have a preference structure in which they cared about

14. See Frohlich, Oppenheimer and Kurki (2004) for mathematical details.

1. Their own payoff;

2. How much less they received than the other person;

3. How much less the other person received than they did;

4. How much less they received than they had produced; and

5. How much less the other person received than they had produced;

The data fit the generalized model quite well as can be seen from the diagram in Figure 8.

There, to display these different modes of rational normative behavior we examine scatter plots comparing the proportion of money left by each dictator with the proportion of the production contributed by that counterpart. In the figure the three sorts of normative behavior are highlighted and labeled. Leaving the 'earnings' of one's counterpart (just deserts) yields a point on the diagonal line; splitting the amount 'equally' falls in the horizontal region half way up the graph; naked self-interest (or leaving nothing) results in an observation on the horizontal axis.

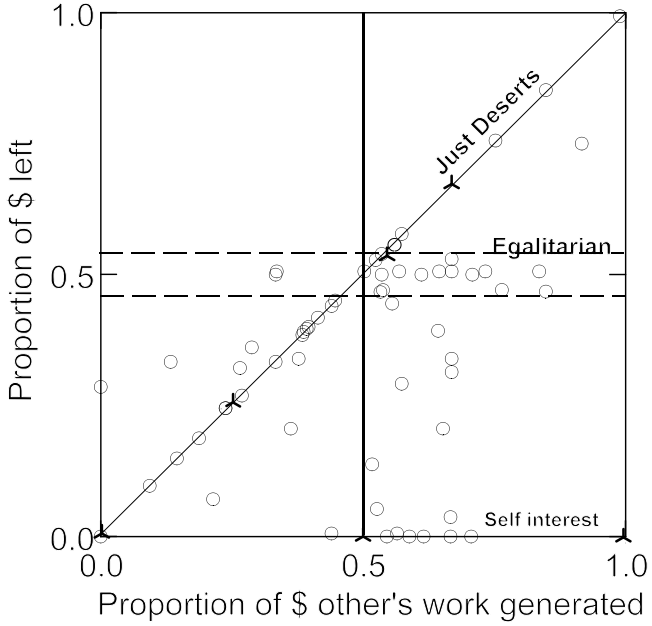

Figure 8: Proportion left with 3 Alternative Normative Principles

The results show three sizeable clusters of individuals: one cluster in each of the of the 3 identified zones. The Fehr/Schmidt model (motivation for avoiding inequality) and our extended model can be compared regarding their predictive ability in this context. Because they make different predictions regarding the distribution of observations over some parts of the payoff space. In particular the original Fehr/Schmidt model of inequality aversion implies that a dictator will never leave more for her counterpart than she takes for herself. In other words, there should be no data points above the horizontal zone (but there are). By contrast, our expanded model implies that individuals can leave more than an equal division when the recipient is more productive than the dictator, but *not* when the recipient is less productive: any points above the horizontal zone should appear to the right of the vertical dotted line (they do). Further, our model implies the amount left will never exceed the maximum of an equal split and just deserts. So no points should fall in the upper portion to the left of the just deserts line (there are none)

So the general problem of other regarding behavior is more complicated than simply tapping attitudes towards others. Different moral values are evoked in different contexts. With this perspective, a moral point of view is a structured cognitive frame an individual uses to evaluates aspects of the decision problem that conforms to some minimal criteria for taking into consideration

others' welfare. Different facets of a moral point of view will be evocable by different environmental cues.[15]

So the problem is more than just non-separable preferences. It appears that preferences vary as a function of contextual factors. To develop a viable micro model of decision making, we need some understanding of how individuals come to interpret situations they encounter as one context or another

Cognition, Preferences, and Rationality

Of the two problems in the standard model: non-separable preferences and context-dependent preferences, the former is, by far, the easier to address. Various attempts (cited earlier) at building utility functions that take into account others' utilities *in specific contexts* have been advanced. The problem of a preference structure that is variable as a function of framing or context effects is much more problematic. While non-separable preferences might be addressable by introducing new fixed elements into a standard utility function, that won't do for context dependent preferences. A model that is to cover different contexts would have to cover how the changes in perspective occur so as to allow for context specific rational choice. We turn to a rough sketch of what such a foundation may look like.

One possible explanation comes from considering what we, as individuals, need for our everyday existence. To survive and thrive, one has to anticipate what is likely to happen in the situations one encounters. To do this one must have some implicit theories of how the world works in these situations. Each individual must have at her disposal a repertoire of representations to make sense of the situations she faces. The existence of that repertoire raises the possibility of more than one representation for any given encounter with the environment. That may account for how framing can affect choices and how framing effects can be so easily evoked.

The Multiple Representation Problem and Points of View:

What can we expect to happen when the individual has multiple representations? Consider the Necker cube in Figure 9 (Gaetz et al. 1998). How many representations of it can we actively hold in our mind at once? Careful studies of perception indicate that the mental representation of a perceived object at any instant appears to be unique even though we may be aware of the possible ambiguity of any given representation. You probably see the cube in one of two ways: with the front face pointing up and to the left of the page or, with the front face pointing down and to the right. But only one of those representations is apparent at any one time.

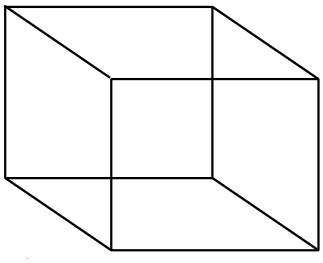

Figure 9: A Necker Cube can be perceived as seen from either above, or from below, but not both simultaneously.

With the Necker Cube, which way one sees it is probably stochastic. In making sense of more complex social situations, in which very different representations are possible, some underlying factors may give the initial edge to one representation over another. Recent or repeated prior use of a representation or specific cues that are linked to particular representations all may play roles in advantaging the call up of one representation over another. How these might interact to explain which representation

15. Roughly, a moral point of view must take into account the welfare claims of others. See Frohlich and Oppenheimer, 2001.)

is actually evoked and how that can affect decisions is a potentially complex story that we can only refer to here (but see Frohlich and Oppenheimer, 2005).

In thinking about how different representations may be called up in trying to make sense of a situation we can rely on an observation of Damasio (1999, 163-64). He describes the process whereby affect is stored and associated with previously perceived objects:

> . . .(T)he memory of .. (an) object has been stored in a dispositional form. Dispositions are records which are dormant and implicit rather than active and explicit, as images are. Those dispositional memories of an object that was once actually perceived include not only records of the sensory aspects of the object, such as the color, shape, or sound, but also records of the motor adjustments that necessarily accompanied the gathering of the sensory signals; moreover the memories also contain records of the obligate emotional reaction to the object. As a consequence, when we recall an object, . . . we recall not just sensory characteristics of an actual object but the past ***reactions*** (emphasis added) of the organism to that object.

But "object" can be extended to include classes or types of situations: competitive games, cooperative games, social banter, care giving, etc. Based on Damasio's observation, each remembered (i.e. recognized) "object" or class of situations has a number of 'emotional reaction' tags associated with it, which can be called up in a situation resembling a relevant class. Different individuals have different life experiences that lead them to react differently to similar contexts. Hence, we might expect a differentiation of individual responses to any given environment (for example, a VCM) – perhaps quite like the variety of 'types' we portrayed in Table 1 and Figures 2-5.

One way of applying this insight is to see how it might "explain" the changes observed in the behavior of subjects in the two types of dictator games described above (Frohlich, Oppenheimer and Kurki, 2004). A traditional anonymous dictator game involves no cues that evoke norms of cooperation. From the dictator's point of view, one's "partner" is anonymous, unknown, *and stands in no recognizable social relationship to oneself.* It is therefore not surprising that no strong norms of sharing are demonstrated in that type of experimental situation. However, when the money to be divided is produced by both participants' efforts, the situation resembles "work." And in work contexts, there is a strong association between working and getting paid. Hence, the situation may be interpreted as including normative elements of entitlements, and dictators may be expected to align the money they leave with their partner's "earned" entitlements.

Another way to apply this insight is to attempt to understand the different types of individuals and their apparent reactions in the VCM experiments described above. First note, that the different levels of disposition to co-operate (in the different types of individuals) might well be explained by the individual's prior experiences in situations resembling VCM's. Her "archetypical" representation of a VCM-like situation might be associated with prior experiences that dispose her, initially, to react in characteristic ways. But how do we explain the unstable behavior we showed in Figures 2-4?

Any individual understands her situation via a representation of it that bring forth a specific preference structure tied to the representation and hence predisposition to cooperate. But recognizing the situation as of a certain type also brings forth a specific sort of monitoring for sensory input regarding her judgement of reality as really being that type. She is looking for specific sorts of cues relevant to optimizing her behavior to the situation (e.g. non cooperation by others). She is constantly implicitly asking herself: "What is going on?" and representations are modified as she gets information that updates the situation. So, in a VCM, with little information and no communication, she might find considerable instability of her behavior given the representation and

as the group behavior fluctuates. "This is a really co-operative group." "No, someone here is really nasty?" Or maybe: "The others don't realize how well we could do if we all co-operated." Her choices might be expected to change as more information is received and her understanding of the situation fluctuates. The situation she faces has parallels in the Necker Cube and many well-known figure/ground optical illusions. Individuals looking at these constructs have their perceptions fluctuate between seeing one image and then another. Perhaps the unexplained saw-toothed behavior is a combination of searching for alternative representations based on cues and a quasi-random switching of perspectives.

Of course, that description is not a full explanation: rather it is just an explanatory sketch. The complexity of the cognitive processes implicit in this view of decision making is potentially daunting, and it poses substantial problems for the theorist.

Some Conclusions and New Directions

We can only begin to sketch what a reconstructed version of rational choice theory would look like. The theory, at this point, is not developed. Yet some of its outlines are beginning to take shape. The first point to note is that the assumption of self-interest, along with its implied inter-personal separability of preferences, can no longer, credibly, be posited as universally applicable to all choice situations. There are simply too many findings, in too many contexts, which do not support the assumption. That is *not* to say that the self-interest assumption fails *universally*. There are contexts, such as markets, competitive elections, military tactics, etc. in which, it can perform admirably. But it is likely destined to play a more limited role in rational choice theory devoted to explaining non-market decisions.

The theory of choice outside of such specific contexts will likely have to make room for a variety of other regarding (hence, non-separable, or non-self-interested) preference structures. We have identified a few alternative elements that are likely to populate preference functions: They include, but are not restricted to, utility (benefits or costs) associated with the size (or possibly relative size) of others' payoffs; the divergence of one's own and others' payoffs from some norms (such as just desert, entitlement, fairness or reciprocity); and a disposition to punish "cheaters."

This list is in the spirit of Ostrom's (1998) Presidential Address in which she addresses the problem of explaining choice behavior with rational choice theory. She notes a number of problems with the traditional rational choice model including the existence of quite arbitrary structural boundaries in the applicability of the models, explicit learning, other regarding behavior, and inconstancy of revealed preferences. Motivated by many of the same concerns we have, she notes that most experiments have found initial level of cooperation significantly higher than Nash would predict; that the general ideas of backward induction, or learning Nash play, has been challenged by failure in a number of major contexts such as VCM's. Ostrom then proposes an effort to develop a more general theory going beyond rational choice theory to cover more of human behavior. She proposes a solution that builds upon the core rationality model by adding a number of structural and preference elements into the tool kit. People would be modeled as complex, fallible learners who explicitly inherit a strong capacity to learn reciprocity norms and social rules that enhance the opportunities to gain benefits (with emphasis on learning over a life-long sequence).

Yet our efforts diverge in a few important aspects. We would argue that structures (such as markets, legislatures, elections, etc.) can be subsumed under a broader notion: context. In context, structure is one, perhaps principal component, and, when 'recognized' by the actor, it affects the

elements that will enter into the individual's decision process. Context must be understood as a cognitively malleable construct susceptible to cueing and framing effects. The result is that many of the elements of the individual's evoked utility function are conditionalized, not only on the behavior of others, but also on the subjectively interpreted context of the decision.[16] In a forthcoming book (Ostrom, 2005, Chapter 4) adopts some aspects of our cognitive model (Frohlich and Oppenheimer, 2005) as possibly contributing to that effort.

Clearly, which utility elements enter, and how they enter for any *objective* context is variable across individuals and is, moreover, susceptible to variation within the individual's experience as a function of how that *objective* context is presented to the individual. Different framings may evoke different utility elements with different strengths. These layers of complexity allow one to use arguments of the form we have presented above to explain seemingly erratic or anomalous behavior, across individuals, and within individuals over time.

But the potential explanatory power of our proposed cognitive - rational choice theory is achieved at a significant cost, bringing to the forefront, epistemological issues of the sort highlighted by MacDonald, 2003. If preferences are context dependent, then traditional rational choice theory's domain is radically narrowed. By introducing additional levels of complexity to preferences, we make the task of falsification more difficult. It may appear to require an instrumentalist justification of each decision model within its particular decision context(s). But such a reading would be wrong. We are proposing a generalization of the rationality argument that knits together decision contexts, and that can be tested (Frohlich and Oppenheimer, 2005). In that sense, the proposal can be improved to decrease our field's reliance on purely instrumentalist based justifications for context limited models. Again, using MacDonald's lines of analysis, our proposal is away from instrumentalism inasmuch as it posits *more realistic, not less realistic* assumptions. It rejects the instrumentalist tack of those who would cling to the projection of models using consistent preferences, or self-interest, in contexts in which those assumptions are not accurate.

On the other hand, our approach does not throw out the baby with the bath water. It leaves open the possibility of modeling existing results which *do* follow from the standard model when the traditional micro premises hold. As Satz and Ferejohn (1994) suggest, this is likely to be in those cases when the structure of the context is so definitive, that it induces stable, and predictably structured preferences and behavior. Their notion of structurally induced interests work, presumably, only when structures are strong enough to crowd out other preference possibilities. Plott (1983) would say that the power comes from the constraint of the choices by institutions. But

---

16. Perhaps presenting a possible mathematical representation of our proposal will facilitate understanding and debate. The expanded utility function of a decision maker in a non-separable, context dependent, version of a rational choice theory might look like the following:

$$U_i(x \mid C_y) = x + \Sigma^*_{it}(C_y)\cdot\alpha_{it}(C_y)\cdot U_{it}(x)$$

Here the utility of a payoff to an individual *i* is a function of a payoff, x, plus a sum of terms $^*_{it}(C_y)\cdot\alpha_{it}(C_y)\cdot U_{it}(x)$, which is comprised of the dummy variable $^*_{it}$ which takes on the value of either zero or one depending on the context, as perceived by the individual, multiplied by the numerical strength of that term, α_{it} (which is also a function of the context (C_y)) and the substantive component of the term, $U_{it}(x)$, and a variety of similar terms. Each "$U_{it}(x)$" term stands for a different element of (likely) non-separable preference. So, one may stand for the effect of a difference between *i*'s payoff and the payoff of others, another for the value to *i* of punishing a cheater, etc. The context dependence of the preferences is represented by both the * term, which determines whether it comes in for a given context C, and the α term which determines the extent to which that utility element impacts total utility for *i*, when it enters in that context.

note, that although the constraints on choices are considerable in a VCM experiment, the behavior is not consistent. Perhaps this is because the VCM, though constraining, is not competitively forcing the actors to follow a narrow interpretation of their self interest. Satz and Ferejohn's notions seem to require that competitive pressure filter out the "noise" of non-separable preferences. But we political scientists are often interested in contexts that do not contain those preference constraining structures. And to apply our thinking in *those* domains requires some micro theory that accounts for the changes in the preferences actors appear to experience with changes in contexts. In our model, contexts (such as a market or a legislative vote) that has been experienced often, and that presents clear incentives to the individual, may evoke narrowly self-interested values and exclude the many complicating factors which can come to play in the VCM. In that sense, the model is a hybrid of a preference driven, and structurally determined model. Yet, at base, there is a presumption that preferences are real, reside in the individual, and are most proximally determinative of the individual's choices.

Satz and Ferejohn's notion of structurally induced interests is used to reject rational choice as an internal psychological theory. Such a move precludes evidence from cognitive psychology as relevant to tests of a rational choice theory of politics. But that move is a step back from the bolder hopes of scientific theory construction wherein implications of basic premises permit additional indirect tests of those premises (also see Hausman, 1995, who criticizes this move). Theirs is an instrumentalist specification of some contexts as the ones in which one can justifiably apply the predictions of the theory. Our proposal, in contrast to theirs, tries to preserve the theoretical core of the rational choice paradigm, as based on internal motivations, by incorporating external structures as generators of context which impinge on and evoke preferences. This is in keeping with the notion of developing a progressive research program (Lakatos, 1970).

There is yet another sense in which this model of non-separable, context dependent, preferences steers a middle course. It identifies a path between models based on methodological individualism and those based on structures. The non-separability of preferences allows one to build in a compromise to the overly demanding self-interest assumption which implies additive separability. Rather than assuming preferences which exclude all concern for others' welfare, save instrumentally, (the preferences of a psychopath), an individual's preferences are tied to a web of social interdependency. The analysis may start at the level of the individual and his preferences, but the structurally cued non-separability of those preferences implies that social structures and norms play a role in the ultimate choices.

Policy Implications

The cognitive approach we are proposing evokes two general implications for classical economic and political theory. First, the standard rationality choice model, simple, elegant and decisive, is not liable to be replaced with as simple and manipulable a model. If we are to maintain a progressive research program that protects the theoretical core, rather than a degenerate one (Lakatos, 1970), we can not erect arbitrary barriers of applicability, as proposed by Satz and Ferejohn. The anomalies that have been identified are too broad and diffuse. The new assumptions are likely to involve a substantially wider array of variables to explain behavior in changing contexts and considerably more theoretical superstructure to address and explain the broad array of phenomena on the table.

Second, the traditional model's view of politics, and especially democracy, as a preference aggregation exercise, is called into question. In the standard model, individuals' unique (and stable and consistent) preferences are posited to be an appropriate basis for evaluating welfare. Most of

welfare economics and public policy evaluation calls for some sort of aggregation of individual preferences based on the presumption of a single underlying preference structure for each individual. But it seems apparent that individuals have non-unique preferences: different preferences over the same choices can be elicited in different contexts, which can be themselves be evoked by different framings. No particular preference structure can have the strong normative claim that is usually attributed to it within rational choice theory. It cannot be argued that any evoked preference is a particularly good 'stable reflection of the welfare of the individual.'[17] The normative justification of Paretian optima, democratic results, and utilitarian analyses based on simple revealed preferences, become suspect: a deeper analysis of group welfare is required. Given our caveat, aggregating expressed preferences would no longer necessarily constitute the best measure of aggregate welfare.

Indeed, the possibility of eliciting different preferences from the same individual by framing contexts, leads to a role for democracy prior to the aggregation task. Politics and political decision making is a matter of relatively low importance to the average citizen in a democracy (given the limited impact a single individual can have, Downs, 1957). And the average citizen does not have a great deal of exposure to making political decisions. Those are the conditions under which we would expect individuals to be maximally susceptible to the differential framing effects. Differential framing of the issues, will elicit different preferences. For example, some appeals may tend to weight or privilege self-interested preferences while other rhetorical approaches can evoke and promote other-regarding preferences. The result of such different framing can, therefore, result in different outcomes and different levels of social welfare under similar material conditions. And, of course, this effect may even be more pronounced between political systems, where both different political structures and appeals by different factions may vary.

From that perspective, this model of rational choice theory gives a role to politics in the "shaping" of preferences, while the traditional model takes preferences as exogenous to the process (if not the structures). They move us from the simplistic notion of 'preference aggregation' to the more fundamental questions that have faced political analysts for years: preference formation and selection.[18] Rather than evaluating democracy only on its ability to generate faithful and meaningful aggregations, one can begin to discuss the problem of the qualities of the preferences being aggregated. It opens up a broader role for the political at both the positive and normative levels. Most promisingly, our approach ought to facilitate new possibilities for integrating the normative and empirical sides of political science.

---

17. This does not imply that individual welfare ought to be considered independently of individual preferences.

18. An interesting essay written from this perspective is contained in Mackie's (2003) pages 99-113.

Bibliography

Adler, Jerry, with Mary Carmichael (2004). "Mind Reading," *Newsweek,* U.S. Edition. July 5, 2004.

Ahn, T. K., Elinor Ostrom, and James M. Walker (2003) "Incorporating Motivational Heterogeneity into Game Theoretic Models of Collective Action" *Public Choice,* special symposium, Essays in Honor of Mancur Olson. Edited by Norman Frohlich and Joe A. Oppenheimer: 295-314.

Andreoni, James. (1995). Cooperation in Public-Goods Experiments: Kindness or Confusion? *American Economic Review*, Vol. 85, No. 4, September 1995. 891-904.

Arrow, Kenneth J. (1977). "Extended Sympathy and the Possibility of Social Choice." *American Economic Review*, 67 (December): 219 - 225. reprinted in Collected Papers of Kenneth J. Arrow, Vol. 1, Social Choice and Justice. Belknap Press of Harvard

Arrow, Kenneth J. (1963). *Social Choice and Individual Values,* 2nd ed. Yale: New Haven.

Ashley, Richard, Sheryl Ball and Catherine Eckel (2004). "Analysis of Public Goods Experiments Using Dynamic Panel Regression Models" (mimeo)

Baumol, William J. and Wallace E. Oates (1979), *Economics Environmental Policy and the Quality of Life.* Prentice Hall, Englewood Cliffs, NJ.

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

Bolton, Gary E. (1991). "A comparative Model of Bargaining: Theory and Evidence." *American Economic Review.* 81 (5, Dec): 1096-1136.

Bolton, Gary E. and Axel Ockenfels (2000), "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review.* 90 (#1, March): 166-193.

Bolton, Gary E. and Rami Zwick (1995) "Anonymity versus Punishment in Ultimatum Bargaining." *Games and Economic Behavior* 10, 95 - 121.

Cain, Michael (1998). "An Experimental Investigation of Motives and Information in the Prisoners' Dilemma Game" *Advances in Group Processes,* 15:133-60, John Skvoretz and Jacek Szmatka editors, New York: JAI Press.

Chaudhuri, Maitra and Graziano, "A Dynamic Analysis of the Evolution of Conventions in a Public Goods Experiment with Intergenerational Advice." (Mimeo: 2004)

Cox, James C. Klarita Sadiraj and Vjollca Sadiraj (2001). "A Theory of Competition and Fairness without Equity Aversion," Paper presented at the International meetings of the Economics Society of America, Barcelona, Spain, June.

Croson, Rachel T.A. (1998). Theories of Commitment, Altruism and Reciprocity: Evidence from Linear Public Goods Games (mimeo).

Damasio, Antonio. (1994) *Descartes' Error.* New York: Avon Books.

Damasio, Antonio. (1999) *The Feeling of What Happens.* New York: Harcourt Brace and Company.

Dickhaut, John, John Hubbard and Kevin McCabe (1995) "Trust, Reciprocity and Interpersonal History: Fool Me Once, Shame on You, Fool Me Twice, Shame on Me." (Mimeo). University of Minnesota.

Eckel, Catherine C. and Philip J. Grossman. (1996) "Altruism in Anonymous Dictator Games." *Games and Economic Behavior.* v. 16: 181-191.

Eckel, Catherine C. and Philip Grossman (1996). "The Relative Price of Fairness: Gender Differences in a Punishment Game." *Journal of Economic Behavior and Organization.* v. 30: 143 - 158.

Fehr, Ernst and Klaus M. Schmidt (1999) "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, Aug.: 817-868

Frank, Bjorn & Gunther Schulze, "Does economics make citizens corrupt," *Journal of Economic Behavior & Organization* Vol. 43 (2000) p 101-113

Frankena, William K. (1983) "Moral - Point - of - View Theories." in *Ethical theory in the Last Quarter of the Twentieth Century.* (Eds. Charles L. Stevenson, William K. Frankena, R. B. Brandt, and A. I. Melden). Indianapolis: Hackett Publishing Company: 39-80.

Frohlich, Norman , Joe A. Oppenheimer, and J. Bernard Moore (2001b) "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 and Joe A. Oppenheimer (2003). "Justice Preferences and the Arrow Problem." Paper presented at the 2003 Annual Meeting of the Public Choice Society, Nashville, TN. March 22-24, 2003. (Available as www.bsos.umd.edu/gvpt/oppenheimer/condwin.pdf)

Frohlich, Norman, Joe Oppenheimer, with Anja Kurki. (2004). "Problems in Modeling Social Preferences: Insights from Modified Dictator Experiments." *Public Choice* 119, Nos 1-2, 91-117. (Available as www.bsos.umd.edu/gvpt/oppenheimer/research/dictprod2.pdf)

Frohlich, Norman & Joe A. Oppenheimer, (2003b), "Optimal Policies and Socially Oriented Behavior: Some Problematic Effects of an Incentive Compatible Device," in *Public Choice.* Special Issue: Essays in Memory of Mancur Olson. V. 117, No. 3-4 (December.): 273-293. (Available as www.bsos.umd.edu/gvpt/oppenheimer/OptPol.pdf)

Frohlich, Norman and Joe Oppenheimer (2001) "Choosing from a Moral Point of View," *Interdisciplinary Economics,* (February) Vol. 12: 89-115. (Available as www.bsos.umd.edu/gvpt/oppenheimer/research/mpv.pdf)

Frohlich, Norman and Joe Oppenheimer (2005) "Choosing." working paper. (Available as www.bsos.umd.edu/gvpt/oppenheimer/research/choosing.pdf)

Frohlich, Norman and Joe A. Oppenheimer (2000) "How People Reason about Ethics." in *Elements of Political Reason: Cognition, Choice and the Bounds of Rationality.* pp. 85 - 107. (eds. Arthur Lupia, Matthew McCubbins, and Sam Popkin). Cambridge University Press. (Available as www.bsos.umd.edu/gvpt/oppenheimer/research/reason.pdf)

Frohlich, Norman and Joe A. Oppenheimer (1998). Examining the Consequences of E-Mail v. Face to Face Communications in Experiments. *Journal of Economic Behavior and Organization.* v. 35: No. 3 (15 Apr. 1998). pp. 389 - 403.

Frohlich, Norman and Joe A. Oppenheimer (1996)."Experiencing Impartiality to Invoke Fairness in the n-PD: Some Experimental Results." *Public Choice,* 86 (117 - 135).

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.

Gaetz, Michael, Hal Weinberg, Edward Rzempoluck, and K.J. Jantzen (1998). Neural Network Classifications and Correlation Analysis of EEG and MEG activity accompanying Spontaneous reversals of the Necker Cube. *Cognitive Brain Research* 6: 335-346.

Grether, David M. and Charles R. Plott, (1979) "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review,* 69 (September): 623 - 638.

Gunnthorsdottir, Anna, Daniel Houser, Kevin Mccabe, and Holly Ameden (2001). Disposition, History and Contributions in Public Goods Experiments. (mimeo).

Hausman, Daniel M. (1995) "Rational Choice and Social Theory: A Comment," *The Journal of Philosophy,* v. 91, #2 (Feb.): 96-102.

Houser, Daniel and Robert Kurzban (2002). "Revisiting Kindness and Confusion in Public Goods Experiments." *American Economic Review*. V92: 4 (Sept.): 1062-1069

Hoffman, Elizabeth, Kevin McCabe, and Vernon Smith. Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology. Mimeo (April, 1996).

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.

Isaac, R. Mark and James M. Walker (1988). Communication and Free-Riding Behavior: The Voluntary Contribution Mechanism. Economic Inquiry, 26 (4) (Oct.): 585-608.

Issac, R. Mark, James M. Walker and Arlington W. Williams, (1994) Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups. Journal of Public Economics, No. 54, 1-36.

Janssen, Marco A. and T.K.Ahn, "Adaption vs. Anticipation in Public-good Games", Sept. 2003 (mimeo).

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1986). Fairness and the Assumptions of Economics. in a specially edited volume (by Robin M. Hogarth and Melvin W. Reder) of The Journal of Business, 59 (No. 4, Part 2: October): S285- S301.

Konow, James. (2000) Fair Shares and Cognitive Dissonance in Allocation Decisions." *American Economic Review*. 90 (4: September): 1072-1091.

Lakatos, I. (1970). "Falsification and the Methodology of Scientific Research Programmes." In *Criticism and the Growth of Knowledge*. I. Lakatos and A. Musgrave (eds.), pp. 91-97. Cambridge: Cambridge University Press.

Larrick, Richard P. and Sally Blount, (1997). "The Claiming Effect: Why Players are More Geneorous in Social Dilemmas Than in Ultimatum Games." *Journal of Personality and Social Psychology,* 72: 810 - 825.

Ledyard, John O. (1995). "Public Goods: A Survey of Experimental Research," in *The Handbook of*

Experimental Economics, John H. Kagel and Alvin E. Roth, eds. Princeton University Press: Princeton. pp. 111 - 194.

Levati, Maria Vittoria. "Explaining Private Provision of Public Goods by Conditional Cooperation - An Evolutionary Approach." Paper given at an NSF sponsored conference at the University of Indiana, Jan. 24-25, 2003.

Lindblom, Charles E. and David K. Cohen. (1979) Usable knowledge: social science and social problem solving. New Haven: Yale University Press.

MacDonald, Paul K. (2003). "Useful Fiction or Miracle Maker: The Competing Epistemological Foundations of Rational Choice Theory." *American Political Science Review,* v. 97 #4 (November): 551-565.

Mackie, Gerry (2003). *Democracy Defended.* Cambridge University Press: Cambridge, England.

May, Kenneth O. (1954) "Intransitivity, Utility, and the Aggregation of Preference Patterns." *Econometrica,* vol. 22 (No. 1, Jan): 1-13.

Olson, Mancur (1965) The Logic of Collective Action. Harvard University Press: Cambridge.

Ostrom, Elinor (2005, forthcoming) *Understanding Institutional Diversity,* Princeton University Press: Princeton, NJ.

Ostrom, Elinor (1998). "A behavioral approach to the rational choice theory of collective action." *American Political Science Review,* 92 (1): 1-22.

Ostrom, Elinor and James Walker . 1991 "Communication in a Commons, Cooperation without external Enforcement," in *Laboratory Research in Political Economy,* Thomas R. Palfrey, ed. 287-322. Ann Arbor: University of Michigan Press.

Ostrom, Elinor, James Walker and Roy Gardner. 1992 "Covenants with and without a Sword: Self-Governance is Possible," *American Political Science Review,* V. 86, No. 2, June, 404-417.

Plott, Charles R. (1983) "Externalities and Corrective Policies in Experimental Markets." *The Economic Journal,* 93 (March): 106 - 127.

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.

Riker, William H. (1982) Liberalism Against Populism: A Confrontation Between the Theory of Democracy and the Theory of Social Choice. Prospect Heights, IL.: Waveland Press.

Rabin, Matthew (1998). "Psychology and Economics," *Journal of Economic Literature,* vol 36 (March): 11 - 46

Rabin, Matthew (1993). "Incorporating Fairness into Game Theory and Economics." *American Economic Review.* v. 83, #5 (December): 1281 - 1302.

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.

Ruffle; Bradley J. (1998) "More is Better, But Fair Is Fair: Tipping in Dictator and Ultimatum Games." *Games and Economic Behavior.* May: 247-265.

Saijo, Tatsuyoshi and Toru Yamaguchi, "The 'Spite Dilemma' in Voluntary Contribution Mechanism Experiments." Paper presented at the 1992 Public Choice Meetings, March: New Orleans.

Satz, Debra and John Ferejohn, (1994) "Rational Choice and Social Theory," *The Journal of Philosophy,* v. 91, #2 (Feb.): 71-87.

Schelling, Thomas C. (1973) "Hockey Helmets, Concealed Weapons, and Daylight Savings: A Study of Binary Choices with Externalities." *Journal of Conflict Resolution,* v. 17, No. 3 (September), pp. 381 -428.

Sen, A.K. (1970) *Collective Choice and Social Welfare.* North Holland: New York.

Sen, A. K. (1973) *On Economic Inequality.* Norton: N.Y.

Sen, Amartya K. (1966). "A Possibility Theorem on Majority Decision." *Econometrica* 34: 491 - 499.

Shafir, Eldar and Amos Tversky (1995). "Decision Making," in Edward E. Smith and Daniel N. Osherson, eds. *An Invitation to Cognitive Science: Thinking* Volume 3, Second Edition, Cambridge, Mass.: MIT Press. 77-100

Shepsle, Kenneth A. and Mark S. Bonchek, (1997) *Analyzing Politics: Rationality, Behavior, and Institutions.* W.W. Norton Publishers: New York.

Simon, Herbert, (1947) *Administrative Behavior.* New York, N.Y.: Free Press.

Simon, Herbert (1992) Bounded Rationality: Discussion, in: *Bounded Rationality and the Cognitive Revolution*; edited by Massimo Egidi and Robin Marris. Brookfield, Vt., USA: Edward Elgar.

Simon, Herbert A. (1986) *Journal of Business,* v. 59, no. 4 pt. 2, pp. S209-S224.

Sternberg, Esther M. (2000) *The Balance Within: The Science Connecting Health and Emotions.* Freeman: New York.

Sweeney, John. (1973). "An Experimental Investigation of the Free-Rider Problem," *Social Science Research.* v. 2: 277-292.

Taylor, John (1996). "Modeling What It Is Like To Be." in Hamneroff, Stuart R., Alfred W. Kaszniak and Alwyn C. Scott, Eds. *Towards a Science of Consciousness: The First Tucson Discussions and Debates.* Cambridge, Mass. MIT Press. 353-376.

Tversky, A. and D. Kahneman (1981) "The Framing of Decisions and the Psychology of Choice," *Science* 221 (Jan. 30): 453-458.

Tversky, Amos and Daniel Kahneman (1986), "Rational Choice and the Framing of Decisions," *Journal of Business,* v. 59, no. 4 pt. 2, pp. s251-s278. Reprinted in Karen Schweers Cook and Margaret Levi, eds. *The Limits of Rationality.* U of Chicago, 1990: 90-131.

Wittman, Donald (forthcoming). "When does Altruism Overcome the Intransitivity of Income Distribution?" in *Rational Foundations of Democratic Politics*, Albert Breton et. al. (eds.) Cambridge University Press.

Zeckhauser, Richard and Elmer Shaefer (1968). "Public Policy and Normative Economic Theory." in Raymond A. Bauer and Kenneth J. Gergen, eds. *The Study of Policy Formation.* Macmillan, The Free Press: New York: 27 - 101.