For publication in a Special Issue of Public Choice, dedicated to the memory and heritage of Mancur Olson.) Norman Frohlich and Joe A. Oppenheimer, eds.) (olsonicdr2sub.wpd)

Submission Draft 2
Comments Welcomed
finished: April 21, 2003

Optimal Policies and Socially Oriented Behavior:
Some Problematic Effects of an Incentive Compatible Device[a]

by
Norman Frohlich
I.H. Asper School of Business, University of Manitoba
Winnipeg, MB R3T 5V4, e-mail: frohlic@ms.umanitoba.ca
and
Joe Oppenheimer
Department of Government & Politics
University of Maryland, College Park, Maryland 20742, USA
e-mail: joppenheimer@gvpt.umd.edu

Abstract: Mancur Olson was pivotal in identifying the formal structure of collective action and the problems of achieving optimal social outcomes with it. Using experimental methods, an incentive compatible device is introduced in a 5-person prisoner's dilemma. The arrangements reflect constructs of Harsanyi and Rawls designed to identify optimal and fair outcomes. The device moves groups towards optimality but its removal negatively affects subsequent behavior, compared to a control with no ICD. This spill-over problem seems to reflect a weakened connection between socially oriented values and behavior, suggesting that ICD's may have unanticipated negative externalities.

a. We would like to thank the Social Sciences and Humanities Research Council of Canada and the National Science Foundation for (grant #01523490) for funding this research, and the Universities of Maryland and Manitoba for their continued support for our work. Avery Cook gave us invaluable help in running the experiments and compiling the data. Also, thanks to Peter Coughlin, Betsy Hoffman, Edy Kaufman, Raymond Lee, Dennis Mueller, Peter Murrell, Mancur Olson, Peter Ordeshook, Tom Schwartz and Piotr Swistak for feedback on earlier drafts. Finally, we thank the faculties of Washington University, who gave interesting and helpful criticisms of an earlier presentation of this work. Earlier versions of this paper were presented at the Annual Meetings of the North American Econometric Society, and at a special meeting of the Public Choice Circle of Japan.

Contents

Figures and Tables

Optimal Policies and Socially Oriented Behavior:

Some Problematic Effects of an Incentive Compatible Device

Introduction: Economic and Philosophic Devices

Contributors to both political philosophy and public finance have traditionally addressed two major questions:

"What are the best (or optimal) outcomes for a given society?"

and

"How can the political and economic institutions of a society be structured to achieve any such optima?"

Members of both disciplines have argued - from Plato and Adam Smith through Olson, Rawls and Friedman - that optima are identifiable and are achievable by the careful construction of social institutions. Economists have focused on the implications of rational self-interested behavior for the achievement of social optima. Olson (1965) presented the first detailed set of guidelines regarding how such optima might be obtained in situations involving social dilemmas, if individuals are rational and self-interested. By contrast, political philosophers have tended to emphasize the importance of ethical or socially oriented behavior in achieving desirable outcomes.

Recently, economists have introduced the concept of an incentive compatible device (hereafter referred to as an 'ICD') as a means of harnessing rational, self-interested behavior to achieve optimal outcomes. (See Clarke, 1971, 1977; Groves, 1973, 1977, Groves and Ledyard, 1977, and Tideman, 1977). The fundamental idea behind those devices is to find 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 group optima. Such devices substitute for - or make redundant - the need for socially oriented or ethical behavior.

This paper reports experimental results illustrating what happens when an ICD is introduced into a collective action problem in a laboratory context. The experiments were designed to shed light on both: how an ICD affects the achievement of socially preferable outcomes in a repeated 5-person prisoners' dilemma (hereafter referred to as a 'PD'), and how the experience of the ICD affects subsequent behavior.

The ICD we utilize derives its structure from arguments of Harsanyi (1953) and Rawls (1971). They introduced a hypothetical device (a veil of ignorance) as an institutional arrangement to generate both fair and optimal outcomes. Their constructs build on impartial reasoning - a mode of inquiry that has long been argued to have ethical significance. We invoke impartial reasoning by implementing a weak "veil of ignorance," and use it as an ICD to allow subjects to identify and to achieve fair and optimal outcomes.

We focus on three questions:

1. Does the ICD of impartial reasoning achieve better outcomes in the PD than regular play?
2. Does the operation of this ICD affect subjects' ethical orientations?
3. Do any of the changes induced by the ICD carry into the future and alter subsequent behavior?

The answers to these questions, in the instance of impartial reasoning, may have implications for the broader question of how incentive compatible institutions affect not only our immediate welfare and behavior, but also our ethical or social motivations and subsequent behavior.

The Choice Situations: The Regular PD and the Impartial Reasoning Incentive Compatible Device

We use a repeated, linear, 5-person PD to explore the effects of an ICD on a social dilemma problem. The salient aspects of PD's, are that each player has a dominant strategy and the choices of those dominant strategies lead to a Pareto - inferior outcome. Hence there is tension between

individual incentives and optimal group outcomes.[1] The experiments consist of two - phases.[2] We examine behavior in the presence of an ICD and then look at subsequent behavior and attitudes after the ICD has been removed. We measure the effects of the ICD on both choice behavior and on orientation toward the social consequences of one's behavior. The behavior observed is compared to that of subjects playing an equivalent number of rounds of the PD in the conventional way. We introduce the ICD to see if it ameliorates free - riding and we also introduce communication as a treatment to compare the efficacy of the ICD with the well known beneficial effects of communication (Ledyard, 1995). We also examine the effects of both the ICD and communication on ethical motivation.

The Regular PD

In our experiments (see Table 1) each individual has a budget of 10 units, and can either keep the 10, or put any proportion of it into a bonus fund.[3] Every unit placed in the bonus fund yields .4 units to each person in the game. Since a contribution of any quantity, **x**, by a player, results in a loss of **x** plus a gain of only .4**x** it is clear that it is individually rational to contribute nothing at all resulting in each player's getting 10 units. However, if each were to give 10 units, all could do better with payoffs of 20 each.

Table 1 about here

Impartial Reasoning as an Incentive Compatible Device

1/ There was no "pre-announced" end point to the experiment since our interest is in the social problem of 'collective action' which rarely have a known end point.

2/ Our design parallels an earlier design of Isaac and Walker, 1988.

3/ For exposition, the tables display only a discrete representation of the game, but it is conceptualized and implemented as a continuous game with strategy choice domain being the closed interval [0,10] and the payoffs [4,26].

The ICD we use in the experiments is structured on notions of impartial reasoning. Philosophers have argued that reasoning impartially supports behavior motivated by ethical concerns. In the PD, this amounts to placing each player in a position which gives equal weight to the interests of all parties.[4] This is accomplished by having each player make a decision while not knowing which of the n players' payoffs she will receive. To accomplish this, each player confronts Table 1, makes a decision, and is then a randomizing device determines which player actually gets which payoff. This ICD changes the incentive structure of the game to those payoffs (expressed as monetary expectations) reflected in Table 2.

To illustrate, imagine a particular player trying to decide what to do under the contingency in which only one other player contributes 10 units. Contributing 10 would mean there were a total of 2 (out of 5) contributors, and 3 non-contributors. *Each* player would then receive an expected payoff consisting of 2 out of 5 chances of being assigned to a position which had contributed and 3 of 5 chances of getting a position which had not. The expected value of those would be:

$$.4(20*.4) + .6(20*.4 + 10) = 14.$$

On the other hand, contributing nothing leaves only one contributor and 4 non-contributors. Under that contingency, *each* individual has a 1 out of 5 chance of getting the contributor's payoffs and a 4 out of 5 chance of getting a non-contributor's payoff. The expected value of that strategy is:

$$.2(10*.4) + .8(10*.4 +10) = 12.$$

Note that the value of not contributing is smaller than the value of contributing. This is true under all contingencies. Thus, the game played under the ICD gives players a dominant strategy of contributing the full 10 rather than nothing. Impartial reasoning generates the optimal outcome. A moment's reflection reveals the underlying behavioral incentives induced by impartial reasoning.

4/ Impartial reasoning in this context is addressed in Frohlich, 1992; and Frohlich and Oppenheimer, 1996a.

Each player shares - ex ante - the fate of every other player in a probabilistic fashion and so must weigh everyone's outcomes evenly. That is the essence of impartial reasoning. If the motivational characteristics of impartial reasoning furnish a foundation for ethical reasoning, then, the resulting outcome can also be claimed to be the fair outcome.

Table 2 about here.

It seems quite natural to assume that most players will play their dominant strategy. Indeed, this is strongly supported by considerable evidence (see for example Isaac et al., 1984 and 1985). By imposing the impartial reasoning ICD we generate decisions identical with that which would come from a purely socially oriented point of view (Frohlich, 1992). It generates a reasonable, fair, preferred, and optimal solution.

Impartial Reasoning and Moral Motivation: To tie this argument more closely into ethical theory, we need some distinctions between better outcomes, (associated with particular actions and their associated end states) and *socially or morally motivated behavior*. Ethical behavior is usually (but not always) defined motivationally and not in terms of conformity to certain external observable attributes. Thus, to evaluate the moral content of behavior we need to get at the subjects' motivations. The conditions of impartial reasoning lead to the optimal and 'fair' choice by eliminating the conflict between self-interest and other-regarding behavior. But it need not affect motivation.

The concerns of political philosophers, and our own, go farther than simply getting better outcomes by using an ICD. Much of the civility of everyday life depends the existence of empathic individual motivations which prevent escalating problems of social dilemmas. These motivations often seem to stem from an underlying sense of fairness possibly involving concern for others'

welfare. This leads us to investigate the effects of the ICD on motivation and the relation between the individual's values and their choices.

In addition we are concerned with another factor central to civil life and also known to affect subjects' behavior in PD experiments: the opportunity to discuss decisions prior to making them (Ledyard, 1995). Discussion in a PD is known to have a profound positive effect on contribution levels. This is relevant to our interests in two ways. Forbidding discussion could lower levels of contributions. It could also inhibit the examination of the ethical content of the situation. If subjects cannot discuss "what is right or fair" it could impede confrontation of ethical aspects of the choice and dilute any "carry over" of behavior or ethical concerns into subsequent play. Consequently, we are interested in examining the role of communication in mediating subjects' behavior and orientations.

Hypotheses:

The considerations above lead us to a number of working hypotheses. First,

Hypothesis 1: *Play of a prisoners' dilemma using the ICD of impartial reasoning will increase individual contributions.*

Next, fairness and concern for others' is a consideration which is likely to enter into individuals' decision making when they behave morally. And most moral theories (see Strang, 1960, for example) would identify full contribution as the fair outcome in this 5-person PD. If ethical concerns are motivators for contribution, and our measures of these concerns are valid, one would expect to find a link between ethical concerns and contributing in the regular plays of the PD.

However, impartial reasoning introduces a complicating factor into a subject's ethical reasoning. The ICD gives her an unequivocal incentive to contribute the maximum amount. This is true whether she is completely selfish or whether she is concerned about others' welfare. Hence, both

ethically motivated and selfishly motivated individuals should be expected to behave the same way. Any increase in ethical motivation would not be detectable by increased contribution levels in Phase 1.

These considerations lead to the following two pronged hypothesis:

Hypothesis 2: *Ethical concerns will have force in explaining contribution levels in regular plays of the PD but will not in the presence of the ICD.*

If, as hypothesized above, playing from an impartial point of view sensitizes subjects to the social content of their decisions, we might expect that their subsequent behavior - in situations in which the ICD is not present - to be affected. It is of interest to see whether any such 'social reorientation' effects occur in subsequent behavior. Thus we explore the following hypothesis:

Hypothesis 3: *Experience with the impartial play of a PD will result in higher levels of contribution in subsequent plays of a normal PD.*

In addition, as in Hypothesis 2, we can conjecture that for those who experienced the ICD, the cognitive and motivational effects of concern for fairness and others' welfare should, after the ICD is removed, be more closely tied to behavior among those that have experienced impartial reasoning in prior plays than among those who have not. That is:

Hypothesis 4: *Concern for fairness and others' welfare will have greater explanatory force in explaining contribution levels in subsequent plays of the PD for those who previously experienced impartial play.*

In addition, we are concerned with the effects of communication, both on levels of cooperation, and as mediating factors in sensitizing individuals to the ethical aspects of their situations. For completeness, we reiterate a well established hypothesis:

Hypothesis 5: *Communication will have significant impact in explaining contribution levels in plays of the PD.*

By observing choices in contexts with and without communication we are able to examine the impact of this variable as well.

Research Design:

A two - phase experimental design permits us to introduce a variety of conditions in Phase 1 and then remove them to compare both the relative effects of the treatments and their legacies. The experiments have two basic treatments and two phases (see Table 3). The treatment conditions are method of play: regular PD and play with the ICD; and with and without communication. Phase 1 of all treatments consisted of eight rounds of play of a 5-person Prisoners' Dilemma (either with or without the ICD and with or without communication). Phase 2 of all treatments was the same: seven rounds of a regular 5-person Prisoners' Dilemma without communication. Eleven control experiments and ten impartial play experiments were run. Subjects did not know how many rounds of the experiment would be run. Group membership was constant throughout the rounds.

In each control, groups of five players were introduced to a five-person Prisoners' Dilemma of the form sketched above. Each subject was seated at a computer and the instructions appearing on the screens for play of the game were read to them.[5] They saw that they could explore the implications of different strategic choices available to them on a built-in worksheet screen available in the software (Oppenheimer, et al, 1987) and then could make a decision.

In Phase 1 of the control treatment (8 rounds) the subjects were informed that they would receive the payoff associated with their own decisions and the aggregate group decisions. But they were not aware of how many times the game would be iterated.

The procedure for the groups with an ICD was similar except that subjects were informed that, for each round, after all players had entered decisions in the computers at which they were seated, a

5/ Full instructions are available from the authors.

random drawing of computer numbers would reassign them to one of the five computers. They would then receive the payoff associated with the decision made by the former occupant of that computer.[6] The decision screens used in this treatment showed the same game as used in the control experiments. Hence, the impartiality was not induced by presenting subjects with a transformed payoff matrix such as that in Table 2. The worksheet and decision they made was based on Table 1 as was that of the control group. This was designed to force them to think through the implications of changing positions. Impartial play was accomplished by the random assignment after all subjects entered their decisions.

Phase 1 of the experiments were run both with and without discussion (see Table 3). In the discussion experiments subjects were allowed to discuss what they wished to do until they felt there was nothing more to be gained from the discussion. We ran both the control and the ICD treatments with and without communication among subjects.

Phase 2 of the experiment was identical for all groups: seven identically administered rounds of a regular 5-person PD without communication. After Phase 1, the impartial groups were told that subsequent rounds would no longer involve random reassignment and that they might take a few minutes to re-examine the worksheet. The control groups were told that a brief pause was required after eight rounds and that they might take a few minutes to re-examine the worksheet. The design is summarized in Table 3.

Table 3 about here

Table 3 also shows the number of times the experiment was run under each condition. Our data set is made up of 21 experimental runs, 105 subjects, and 1785 decisions.

6/ Payoffs were 'advertised' in francs. The conversion rate was 17 to a dollar.

At the end of the experiment (that is, after all 15 rounds had taken place) a questionnaire was administered to solicit information about the subjects and their attitudes to test for relationships between choices and concern for fairness and others' welfare. Subjects were asked the extent to which they agreed (retrospectively) that two factors - fairness and concern about others' payoffs - were important in their choices in each of the two phases of the experiment. These questions were: ..(In the) first (second) series of choices ...

"Doing my fair share was important to me."

"Concern about the payoffs of others was important to me."

They were asked to indicate their level of agreement by placing an 'X' on a line of the following sort:

Disagree Strongly |----------| Agree Strongly

Data for testing the four hypotheses in the experiments consist of the levels of contributions of the subjects in the different treatment groups in the two phases and the attitudinal data gathered in the questionnaire.

Experimental Results

Results Phase 1:

Hypothesis 1: *Play of a prisoners' dilemma using the ICD of impartial reasoning will increase individual contributions.*

Hypothesis 5: *Communication will have significant impact in explaining contribution levels in plays of the PD.*

Table 4 about here

Table 4 shows both strong main effects and an interaction effect in the analysis of variance. A large proportion of the variance is explained overall (81.9%) . The communication treatment has a very large independent effect, the ICD also has a strong effect and the two interact. The data in Table 5 detail the somewhat complex story behind the model. There, individuals' average

contribution over Phase 1 are reported. In the absence of communication, the ICD outperforms regular play quite handily: average contributions of 7.42 v. 2.78 . As expected, communication increased contribution levels in both the ICD and regular treatments. However, with communication, contributions in the regular PD treatment were higher than in the ICD treatment by a small but significant amount: 9.99 v. 9.54.

Table 5 about here

But what about *ethically motivated behavior*? For the behavior to be ethically motivated there must be a tie between concern and choice. Evidence bearing directly on Hypothesis 2 must test the *relationship* between levels of ethical concern and contribution levels in the two treatments. To check for this, we test Hypothesis 2.

Hypothesis 2: *Ethical concerns will have force in explaining contribution levels in regular plays of the PD but will not in the presence of the ICD.*

The direct evidence bearing on this hypothesis is obtained by relating subjects' responses regarding their ethical concern to their actual behavior in the two experimental conditions. But, a test of the impact of ethical concern is not straightforward given that communication is such a potent factor in motivating contributions. With communication, both in the ICD and regular PD treatments, there is virtually no variance in contribution levels to explain since the mean contribution levels were 9.54 and 9.99, respectively. Hence the hypothesis is testable only with the data from the treatments without communication.

To test the conjecture, we used answers to the two questions noted above, and constructed an index, which is referred to below as the "level of ethical concern," or "ethics". It consisted of the sum of responses to those two questions. Using only the no communications condition, on a 118

point scale the means (SD) for the two variables were, respectively: (fair share) 66.5 (35.0) and concern for others 49.3 (37.2). The variables were correlated with a Pearson's r of .369 significant at the .01 level and are somewhat skewed G1: -.417 and .202. Summed as an index, on a scale of 236 the mean (SD) was 115.9 (59.7) with a skewedness of -0.424.

The regressions showing the relationships between ethical concern and contributions are presented in Table 6.

Table 6 about here

As the data demonstrate, our measures of ethical concern have a very different impact on cooperative behavior in the ICD and non-ICD treatments. Virtually none ($R^2 = .025$) of the variance is explained by ethical concern in the ICD treatment and the regression is not significant. On the other hand, approximately 36% of the variance in contributions of the normal PD is explained by this measure of ethical concern. The relationship is strongly significant and furnishes support for the hypothesis that those contributions are *ethically motivated*.

Moreover, this lack of explanatory force when there is an ICD holds despite the fact that the level of ethical concern expressed during Phase 1 of the experiment with the ICD was **higher** than that expressed without (130.2 v 101.5). Choosing when there is an ICD, *requires* that the individual calculate and incorporate the effect of the possible decisions on each person's welfare in order to pursue one's self-interest. Only by doing this can she maximize the return to herself. Hence, comparing a person with great ethical motivation with those without any such motivation leads to *no observable behavioral difference*.

Results Phase 2:

Phase 2 of the experiment was designed to explore the residual effects of the Phase 1 experience. In Phase 2 of the experiment, all subjects face the same standard PD scenario: with

neither an ICD nor any communication. Might impartial play have a positive carry-over effect on subsequent contributions? And how would the presence or absence of communication affect any such relationships?

Hypothesis 3: *Experience with the impartial play of a Prisoners' Dilemma will result in higher levels of contribution in subsequent plays of normal Prisoners' Dilemmas*

As the data in Table 7 indicate, there is no significant difference in the levels of cooperation in the treatments without communication. Both ICD and regular PD subjects contributed similar amounts. However, there is a significant difference in the levels contributed in the two treatments with communication. But note that *the difference is in the opposite direction to that posited in the hypothesis.* With communication, playing the PD with the impartial reasoning ICD in Phase 1 leads to *significantly lower levels of contribution* in subsequent normal plays of the game than does playing a regular PD in Phase 1 (2.43 vs. 5.30 units).

Table 7 about here

So there are two notable conclusions from the data. The ICD doesn't lead to higher contributions in Phase 2: *it leads to lower levels of contribution*. But that effect only takes place in the presence of communication.

But now, what is the connection between ethical concerns and contribution levels in Phase 2 of the experiment? We had hypothesized:

Hypothesis 4: *Concern for fairness and others' welfare will have greater explanatory force in explaining contribution levels in subsequent plays of the PD for those who previously experienced impartial play.*

The relevant data can be found in Table 8. As in Phase 1, ethical concern plays a role in explaining contribution levels *for those who played the regular PD.* Both with and without communication there is a substantial relationship between those concerns and contribution levels,

accounting for between 30% and 35% of the variance in the latter. However, contrary to the hypothesis, there is a weaker association between ethical concern and contribution levels in subjects who had previously played the PD with the ICD in place. In the ICD treatment, without communication, there is *no* significant relationship between ethical concern and contribution levels. With communication the explanatory power of expressed ethical concern is roughly half as great among subjects who experienced the ICD in Phase 1 as among those who played a regular PD in that phase. The explained variance is about half as great and the coefficient is less than half the size. Experiencing the ICD in Phase 1 seems to have attenuated the link between ethical concern and ethically motivated behavior.

Table 8 about here

Discussion:

Let us take a step back and survey the results from a broader perspective to get a feel for the main conclusions suggested by the data. Figures 1 to 4, depict, graphically, the individual contributions in the four treatments over the full 15 rounds. Each point represents one individual's contribution in each of the rounds.[7] There are a few broad conclusions that one can draw from the graphs.

In Phase 1 (to the left of the vertical center bar in each of the figures), the patterns in all the figures except in Figure 1 (regular PD, no discussion) are quite similar. Both the ICD and discussion (Figures 2 through 4) generate close to optimal outcomes. But there is a difference, and it is in favor of discussion, not the ICD.

7/ The graphs use a 'distance weighted least squares' smoothing algorithm which permits the line to flex locally. The vertical lines between rounds 8 and 9 divide phases 1 and 2.

In Phase 2 (to the right of the vertical line in each of the figures), again, one condition outperforms all the others. Discussion when subjects play a regular PD in Phase 1, (Figure 3) leads them to do uniformly better in Phase 2 than did subjects in any other treatment. With discussion and regular play the group always does better and is left substantially (and significantly) better off after the 15 repetitions. In that condition the mean contribution level in the last round is 4.2 out of a possible 10, whereas in the other treatments the last round contributions range from between 1.4 to 2.0.

These observations and the analysis presented above allows us to provide tentative answers to the three questions posed at the outset.

First, the imposition of the impartial reasoning ICD generates near optimal outcomes in a small group PD but it does not outperform discussion in that context. Talk outperforms the institutionalization of an ICD in generating individual contributions. This has implications for policy prescriptions for situations such as those commonly portrayed as subject to the "tragedy of the commons" (Hardin, 1968). At least in small groups, discussion is likely to be adequate to generate cooperation, and could outperform an ICD. Of course, this may be much less policy relevant for a larger scale, highly mobile and anomic population, in a social structure in which there is little continual face to face communication. But the finding is certainly relevant for evaluating the simpler arrangements in agrarian situations.

Second, the operation of this ICD affects subjects' ethical orientations substantially. Experience with a regular PD leads subjects to act in a way in which ethical concerns explain a substantial amount of the variance in contributions. For subjects who experience the ICD, ethical concern is either much less effective or totally ineffective in explaining contributing behavior. This is true of subsequent behavior of subjects who play the two types of games as well. Regular PD players show a stronger link between their ethical orientation and behavior than do ICD players.

This last observation bears on our final question: What, if any, behavioral effects of playing under an ICD carry into the future? Our conjecture had been that experience with the impartial reasoning ICD might engender more contributions subsequent to the device's removal. That didn't happen. Without discussion, there was no carryover effect. With discussion, the ICD, possibly due to the decoupling of ethical concern and behavior, led to significantly lowered levels of contributions than were found in the regular PD.

These observations lead us to conclude that the use of an ICD as a policy tool can be a two-edged sword, and ought to be studied further before being advocated widely. It may improve a situation in the short run, and hence remains attractive especially when communication among group members is not practicable. However, the outcomes that may follow after subjects have been exposed to this institution can be worse than might have been obtained in the absence of the ICD.[8] The shortfall has two aspects. It leads to both a degraded outcome, and behavior that is less ethically motivated. This is not what we anticipated and it calls for some discussion and re-evaluation.

One possible explanation for better outcomes in some instances and, at the same time, perversely affected motivation and subsequent behavior is offered by a close consideration of the way in which this incentive compatible device operates on individuals' motivations and perceptions. Consider the choice structure faced by individuals playing a PD from an impartial point of view. With the ICD, individuals confront a situation in which their self-interest and the interests of all others coincide exactly. What is best for them is, by explicit design, best for the group as a whole. There is no tension whatsoever between the best strategy from a rational self-interested point of

---

8/ These results are consistent with some findings in social psychology. When individuals are rewarded for activities which they previously performed for "non-rewarded" reasons, with the stopping of the rewards, the activity level falls below that observed in the initial stages (see Amabile, et. al., 1986, and Lepper, and Green, 1975). More generally, it is believed that the presence of a salient external motivator decreases the *intrinsic interest of the individual* in the activity.

view and the ethically best strategy. Thus, subjects need not take into account the effects of their choices on others *as distinct* from their own calculated self-interest. They can make the calculations solely on a self-interested basis without conflict with other-oriented values. That is, after all, the essence of incentive compatibility. Thus, the implementation of an incentive compatible device actually obviates the need for ethical reasoning. As Steve Turnbull commented: "They don't have to flex their ethical muscles."

This contrasts with the situation faced by subjects who play a regular PD in all rounds. At each stage there is tension between the strategy that is best from a self-interested point of view and the ethically best strategy. Any choice has to take into account competing imperatives. And communication brings to the fore the ethical imperatives relevant to the dilemma.

The differences in the two treatments may follow from the simple observation that people with different ethical motivations will not behave differently in situations in which ethics and rational self-interest coincide. *The impartiality mechanism or any other ICD renders the need to invoke ethical concerns as a motivator moot.* Both ethically motivated and selfishly motivated players can agree on the best strategy when a situation involves an ICD. As a result, when ICD players subsequently have to make ethical decisions they are more likely to downplay the ethical components than are those regular players who have had practice confronting ethical issues.

Another way of looking at these effects is through the lenses of Tversky and Kahneman (1986). They demonstrated that framing choice situations can affect behavior (see Quattrone and Tversky, 1988, for a bibliography). While Tversky and Kahneman focused primarily on losses and gains as the aspects of a situation which can affect choices, they also noted that other attributes of a situation can affect behavior. Of course any situation has an infinite number of possible aspects and individuals must light upon some subset of those aspects to make sense of the situation. The

aspects which individuals focus on either determine or are determinative of different cognitive models which they may use to interpret the situation, to understand the relationships between different aspects of that situation, and to choose courses of action.

One possible inference from our results follows from the observation that individuals are not always sensitive to all aspects of a particular decision which they confront. Their choices are likely governed by which "model" they evoke to make sense of a particular phenomenon. And the model evoked may, in general, be sensitive to the framing of the choice situation. An ICD may "frame" a situation in such a way as to draw attention away from the ethical aspects and to encourage interpretation in terms of naked self-interest. This effect may persist in coloring the interpretation of subsequent situations encountered by individuals. Discussion, especially in regular play, may have the opposite effect: focusing attention on the ethical aspects of the situation. This is evident both in the greater explanatory power of ethical concern in Phase 1 in the Regular PD treatment as well as the higher subsequent contributions and its link to ethical concern in Phase 2.

Experimenters in psychology have long been sensitive to the impact that their framing can have on subjects' behavior in the laboratory. Other social scientists have paid far less attention to the possible framing effects that they invoke when they introduce social interventions like incentive compatible devices as public policy structures. If our interpretations are correct, they raise a fundamental question about an unanticipated externality of incentive compatible devices. There may be an explicit tension between the use of incentive compatible devices and ethical behavior.

To the extent that our goals are both improving collective welfare patterns *and* the fostering of ethical individual responsibility we must concern ourselves with the difficulty of combining the two in a single set of formal institutions. Although discussion (as has been argued by virtually all proponents of democracy) appears to be uniformly beneficial in the examined contexts, the two

goals of optimal outcomes and socially motivated behavior could be difficult to achieve simultaneously via an incentive compatible device.

Bibliography

Amabile, T.M., B.A. Hennessey and B.S. Grossman (1986). "Social Influence on Creativity: The Effects of Contracted - for Reward," Journal of Personality and Social Psychology, v. 50, pp. 14-23.

Clarke, Ed. (1971) "Multipart Pricing of Public Goods, Public Choice, Fall, 2: 17-33.

Clarke, Ed. H., (1977) "Some Aspects of the Demand Revealing Process," Public Choice, 29, no 2 - supplement (Spring), 37-51.

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 (1996a). "Experiencing Impartiality to Invoke Fairness in the N-PD." Public Choice. 86, 117-135.

Groves, T. "Incentives in Teams," Econometrica, July 1973: pp. 617 - 631.

Groves, T. and J. Ledyard.. "Some Limitations of Demanding Revealing Processes," in Public Choice, Vol. XXIX-2, Special Supplement to Spring 1977: 107-124.

Groves, T. and J. Ledyard, "Optimal Allocation of Public Goods: A Solution to the Free Rider Problem," Econometrica, (May) 1977, v. 45: 783-809.

Hardin, Garrett (1968). The Tragedy of the Commons. Science, v. 162: 1243-1248.

Harsanyi, J. (1953) "Cardinal Utility in Welfare Economics and the Theory of Risk-Taking," Journal of Political Economy 61: 434-35.

Isaac, R. Mark, Kenneth F. McCue and Charles R. Plott, (1985), "Public Goods Provision in an Experimental Environment." Journal of Public Economics, 26, 51 - 74.

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

Isaac, R. Mark, James M. Walker, Susan H. Thomas, (1984) "Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations." Public Choice, 43, 113-149.

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.

Lepper, M.R. and D.M. Greene (1975). "Turning Play into Work: Effects of Adult Surveillance and Extrinsic Rewards on Children's Intrinsic Motivation." Journal of Personality and Social Psychology, v. 31, 203-210.

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

Oppenheimer, Joe A., Mark Weiner, and Hsiu Lu (1987) "C&C, Conflict and Cooperation: An Authoring System for Simulations" PDS Software, Chevy Chase, Md.

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.

Rawls, John, 1971. A Theory Of Justice, Cambridge: Harvard University Press.

Strang, Colin (1960) "What if Everyone Did That?" Durham University Journal, 53 (1960), pp. 5-10. Reprinted in Baruch A. Brody, ed. Moral Rules and Particular Circumstances. pp. 135 - 144. Prentice Hall: Englewood Cliffs, New Jersey. 1970.

Tideman, T. Nicolaus, ed. (1977) Public Choice, Vol. 29, no. 2. Special supplement to Spring 1977, on alternative demand-revealing procedures.

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.

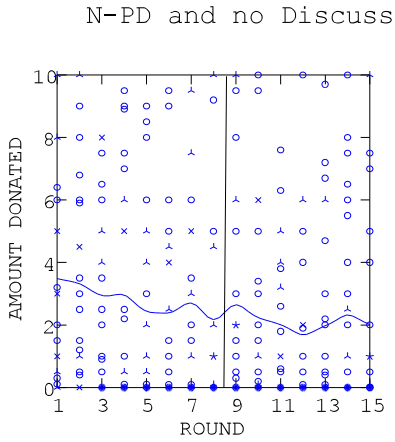

Figure 1 Relationship of Donations to Rounds: Regular PD, Non Discussion Experiments

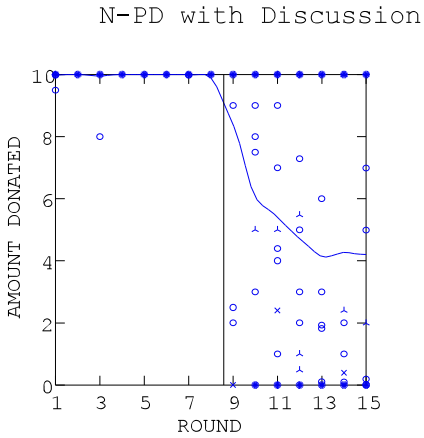

Figure 3: Relationship of Donations to Rounds: Regular PD, Discussion Experiments

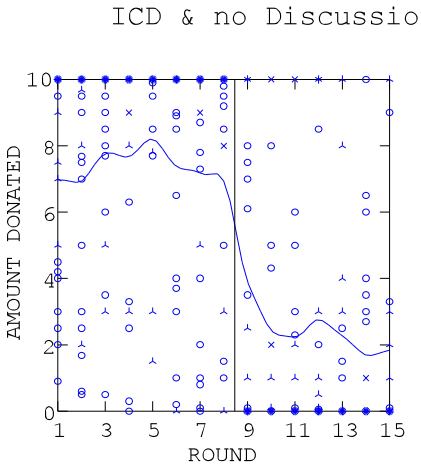

Figure 2 Relationship of Donations to Rounds: Impartial Play, Non Discussion Experiments.

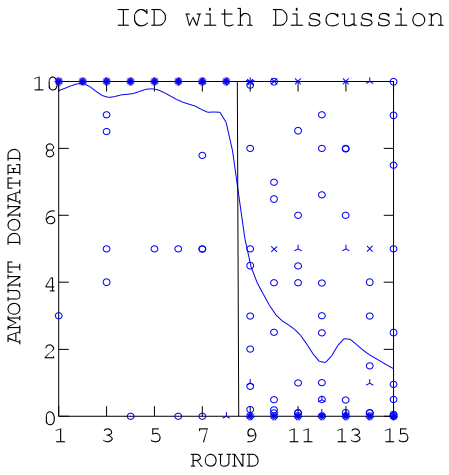

Figure 4: Relationship of Donations to Rounds: Impartial Play, Discussion Experiments

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

Table 2: Impartial Transform of the 5- Person Prisoners' Dilemma (Showing Payoffs to One Player)	Amount Given by Others				
1 Person's Strategies	**40**	**30**	**20**	**10**	**0**
give 0	18	16	14	12	10
give 10	20	18	16	14	12

Table 3: Research Design of the 5-Person Games				
Phases	*Treatments (number of Groups in Each Treatment)*			
	Regular PD		Impartial PD	
Phase 1:	No Communication	Communication	No Communication	Communication
(8 Rounds)	(5)	(6)	(5)	(5)
Phase 2:	Regular PD with No Communication			
(7 Rounds)				

Table 4: Individual Contributions in Phase 1 as a Function of ICD and Communication

Analysis of Variance

Source	Sum of Squares	Degrees of Freedom	F-ratio	P
ICD	114.6	1	61.9	0.001
Communication	567.7	1	306.8	0.001
ICD*Communication	169.5	1	91.6	0.001
Error	186.9	101		

N: 105 Squared multiple R: 0.819

Table 5: Individual Contributions in Phase 1 by ICD and Communication

	No Communication			**Communication**		
Experimental Treatment	**Mean Given**	**SD**	**N**	**Mean Given**	**SD**	**N**
ICD	7.42	1.57	25	9.54	0.649	25
Regular PD	2.78	2.22	25	9.99	0.047	30
Significance: (Mann-Whitney U Test) statistic =	593.50 (p = .0005); Chi-square approximation = 29.732 with 1 df			239.0 (p = .003); Chi-square approximation = 8.59 with 1 df		

Table 6: Contribution Levels as Explained by Ethical Concern in Phase 1 -- No Communication (*** p < .0005)		
Regression	**Model 1: Impartial PD**	**Model 2: Regular PD**
Constant	8.38***	.435
	(.841)	(.676)
Ethical Concern	-.006	.022***
	(.006)	(.006)
Explained Variance	r^2=.025	r^2=.365
***** p < .0001**	n=24	n=24

Table 7: Individuals' Contributions in Phase 2 by Treatments						
Experimental	**No Communication (Phase 1)**			**Communication (Phase 1)**		
Treatment	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**
ICD (Phase 1)	2.41	2.39	24	2.43	2.46	25
Regular PD (Phase 1)	2.14	2.22	25	5.3	3.85	30
Significance: (Mann-Whitney U Test) statistic =	307.0 (p = .889); Chi-square approximation = 0.020 with 1 df			200.5 (p = .003); Chi-square approximation = 8.759 with 1 df		

Table 8: Contribution Levels in Phase 2 as Explained by Ethical Concern in Phase 2 (* p < .0005, ** p < .001)**

	No Communication		Communication	
	Model 1: Impartial PD	**Model 2: Regular PD**	**Model 3: Impartial PD**	**Model 4: Regular PD**
Constant	1.546	.056	.709	-0.571
	(899)	.689	(.831)	(1.700)
Ethical	.008	.023**	.016*	.037**
Concern	(.008)	(.006)	(.006)	(.010)
Significance	r^2=.005	r^2=.355	r^2=.173	r^2=.302
	n=23	n=24	n=25	n=30
Mean Final Round Contribution	1.85	2.008	1.42	4.204