BURCH CENTER WORKING PAPER SERIES

University of California, Berkeley
Department of Economics - #3880
Berkeley, CA 94720-3880

How Strategy Sensitive are Contributions? A Test of Six Hypotheses in a Two-person Dilemma Game

Gary E Bolton
Pennsylvania State University

Jordi Brandts
Institut d'Anàlisi Econòmica, Barcelona, Spain

Elena Katok[12]

Pennsylvania State University

March 1999
Burch Working Paper No. **B99-06**

ɛial support for this project was provided by the Spanish DGCICYT (PB92-0593, PB93-0679 and PB94-0663-C03-01). This papɛ
ɔmpleted during a stay at the Department at the University of California, Berkeley as a visiting scholar. Financial support provided
 Fellowships Program is gratefully acknowledged, as is additional support from the Ministerio de Educación of Spain, DGICYT F
ıd CICYT SEC96-2300. This research was initiated while Bolton and Katok were visitors at the Institut d'Anàlisi Econòmica. The
lly ackowledge the Institute's financial support.

Abstract

We test six hypotheses for contributions in dilemma games, a category that includes the prisoner's dilemma and public goods games. The hypotheses we consider are taken from a wide range of models, and each can explain some aspect of contributing behavior observed in previous studies. Our experiment focuses specifically on the strategic interdependence of contributing behavior, and manipulates the strategy space of a two-person dilemma game especially designed for the task. The hypothesis that contributors have non-linear preferences over own and the other player's payoffs accurately matches the strategic pattern of contributing that we observe across treatments. None of the reasons for contributing advanced by the other hypotheses, whether alone or in additive combination, does so.

Gary E. Bolton and Elena Katok
Department of Management Science and Information Systems
303 Beam
Penn State University
University Park, PA 16802, USA
geb3@psu.edu, exk106@psu.edu

Jordi Brandts
Institut d'Anàlisi Econòmica (CSIC)
Campus UAB
08193 Bellaterra, Barcelona, Spain
brandts@cc.uab.es.

1. Introduction: Explaining the strategy sensitivity of contributions

While some people free ride on the public good, others contribute to it, and do so in amounts that are meaningful. People contribute time and money enough to sustain charities, and vote often enough to sustain democracy. Economists studying public goods games in the laboratory observe much the same: A significant number of people contribute to the public good, even as others free ride.

The standard conceptual vehicles for studying these issues are *dilemma games*, a class of stylized strategic encounters that encompasses the prisoner's dilemma, public goods and common pool resource games, among others. All dilemma games share two defining characteristics. First, if players are rational and self-interested, then free riding is a dominant strategy, producing the highest possible individual payoff regardless of what others do. Second, deviation from the dominant strategy leads to a higher joint payoff for the group, and enough contributions produce an outcome Pareto superior to the dominant strategy equilibrium outcome.

Rational self-interest provides a clear and convincing explanation for free riding. But how do we explain the contributing? One can seek an explanation on two different levels. On one level, one can ask whether contributions are, perhaps in some disguised way, consistent with standard notions of individual self-interest. Alternatively, we can accept the propensity to contribute as fact, and then ask 'What characterization of contributing behavior suffices for the purpose of accurately predicting dilemma game outcomes?' The latter approach has a long tradition in economics (ex., Hochman and Rodgers, 1969; Becker, 1974), and it is the approach addressed by the experiment we describe here.

The right way to characterize contributing behavior is a matter of some controversy, with many competing hypotheses. Dawes and Thaler (1988) organize their survey around two core concepts. *Altruism* posits that contributions are an expression of concern for the payoff of other people. *Norms of cooperation* posits that people play by rules that are inconsistent with standard assumptions about strategic thinking. Ledyard (1995), in a review of public goods experiments, adds a third category, *mistakes*: "[S]ubjects make mistakes, do not care, are bored, and choose their allocations randomly," p. 170.

Each of these categories – altruism, norms of cooperation and mistakes – encompasses a number of specific hypotheses. It is not practical to consider all of them at the same time. Instead, we focus on two prominent hypotheses from each category. Each hypothesis we consider is derived from a model of contributing behavior found in the literature, and each can explain some aspect of contributing behavior identified in empirical studies. We add that our experiment is very simple, and the reader should find it straightforward to take any omitted explanation to the data.[3]

Our experiment compares hypotheses on the basis of *strategy sensitivity*, the extent to which a player's contribution depends on beliefs about others' contributions. Ledyard's statement of the mistakes hypothesis, for example, is strategy *in*sensitive. For each hypothesis-category, we examine one hypothesis that is strategy sensitive and one that is not. The strategy sensitive hypotheses differ on how contributions are predicted to shift with information about others' contributions.

The experimental treatments manipulate the strategy space of a simple two-person dilemma game, with the consequence of changing the information available to a player about the other player's contribution. It is conceivable that reasons for contributing change as the number of players increases. We return to comment on this issue later; for now we note that the experiment's design can, in principle, be extended to games of more than two persons. Hence the present two-player study provides a baseline for studying this issue further.[4]

We emphasize that our experiment does not rule out any given hypothesis as the explanation for *some* individual behavior. What we do find is that only one of the hypotheses correctly tracks the pattern of shifts and non-shifts in overall contributing that we observe across treatments. None of the other hypotheses, whether considered alone or in additive combination, does so. The significance of this result has to do with strategic model construction. In particular, the result is important information about what a model must do to track some basic features of what we will see to be a strong strategic interdependence in player contributions.[5]

---

[3] Including the additional hypotheses that we are aware of would not change our conclusions.

[4] Isaac and Walker (1988) provide experimental evidence that contributing behavior in public goods games may be sensitive to the number of players.

[5] Studies such as Offerman, Sonnemans and Schram (1996) ask 'What kinds of behavior are there and what is the proportion of each in the population?' whereas we are interested in the relative predictive power of the various kinds of behavior. The two lines of research are complementary.

2. Hypotheses from a range of models

We describe the hypotheses in the context of the domain to be investigated. Our experiment concerns the 2-player payoff matrix displayed in Figure 1.

Figure 1. Dilemma Game Payoff Matrix (payoffs in Spanish pesetas)

		Options available to A					
		1	*2*	*3*	*4*	*5*	*6*
Options available to B	*2*	A gets 1750 B gets 500	A gets 1700 B gets 700	A gets 1650 B gets 900	A gets 1600 B gets 1100	A gets 1550 B gets 1300	A gets 1500 B gets 1500
	1	A gets 750 B gets 750	A gets 700 B gets 950	A gets 650 B gets 1150	A gets 600 B gets 1350	A gets 550 B gets 1550	A gets 500 B gets 1750

The matrix reflects a particular linear public goods payoff function: Player *A* chooses between six levels of contribution, $g^A = 1,...,6$. For each contribution level above *1*, *A* incurs a marginal cost of 50 pesetas and *B* realizes a marginal benefit of 200 pesetas. Player *B* chooses between two levels of contribution, $g^B = 1$ or *2*. Contributing *2* costs *B* 250 pesetas, and benefits *A* 1000 pesetas. In our experiment, the payoff matrix is always common knowledge to the players. The standard dominant strategy argument implies that each player contributes the minimum, *1*, independent of what either player knows of the other's choice. (The dominant strategy argument is therefore strategy *in*sensitive.)

By design, the primary focus of the analysis is on *A*'s choice. To make the game as cognitively simple for *A* as possible, we restrict *B* to two choices: *B* either contributes or he doesn't, there are no subtle shades. On the other hand, giving *A* six choices, rather than say two, provides a detailed portrait of *A*'s motivations.[6]

---

[6]To see the point, suppose we restricted *A* to just two of his choices, say *1* and *6*. Suppose further that *A* knows that *B* chooses *2*, and that under this condition we observe all *A*'s playing *1*. We might be tempted to take this as evidence for pure self-interest, but in fact it only shows that *A* is not so concerned with B's payoffs as to give up 1250 for *B*. We cannot rule out that, had *A* been given the opportunity to play *3*, he would have given the more modest gift of 100.

Dominant strategy equilibrium does not admit contributing behavior. On the other hand, it is a simple, well-defined argument, and one that we find provides a useful organizing framework for posing the hypotheses we want to test. We express the dominant strategy argument in terms of three components:

C1. Player *A* prefers a higher value of his payoff, x^A, to a lower value.

C2. Player *A* can identify the action that best satisfies C1.

C3. Player *A* takes the action that he believes best satisfies C1.

The dominant strategy argument is assembled as follows: Player *A* sees that $g^A = 1$ maximizes the amount of money he will make no matter what the other player does (C1 and C2). He chooses $g^A = 1$ (C3).

Each component, C1 through C3, represents a well-defined conceptual building block: payoff preferences (C1), cognition (C2) and choice (C3). Roughly speaking, each hypothesis we test varies one component while holding the other two fixed. Altruistic hypotheses modify C1, cooperative norms modify C3, and mistakes modify C2.

We now lay out the hypotheses using the payoff matrix in Figure 1 to clarify some of the ideas. For each category (altruism, norms of cooperation, mistakes) there are two hypotheses; one that is strategy sensitive and one that is not.

2.1 Altruism Hypotheses: Alternatives to C1

Altruism posits that people have preferences over the other player's payoff as well as their own. Hence the two altruism hypotheses reformulate payoff preferences.

Linear altruism (C1.a). Contributing player *A* has preferences over the commodity bundle given by $u(x^A, x^B) = x^A + ax^B$, where $a > 0$ is the coefficient of altruism.

This straightforward way of modeling altruism has a long tradition in economics, going back to Edgeworth (1881). It was recently used by Anderson, Goeree and Holt (1998) as part of the explanation for linear and quadratic public goods games with more than two players.[7]

In a public goods environment like ours linear altruism makes an all or nothing prediction: Player *A* will contribute either 1 or 6. The next hypothesis allows for intermediate behavior.

Distributive preferences (C1.b): Contributing player *A* has non-linear preferences over (x^A, x^B) such that A's contribution will be larger when B chooses *2* than when B chooses *1*.

Several specific models imply the distributive preferences hypothesis. For one, there is Becker's (1974) seminal analysis of the public goods problem. One can formulate A's decision problem as choosing a bundle (x^A, x^B) subject to $x^A+1/4x^B= w^A+1/4w^B$. The right hand side of this equation is what Becker calls social income. We now conceive of the payoffs in column 1 as the amount of available social income, which is larger in row 2 than in row 1. Assuming that both arguments in the utility function represent normal goods (something we can test the data for), then as the available social income rises, so does *A*'s choice of both x^A and x^B.

Another model that implies the distributive preferences hypothesis is the ERC model of Bolton and Ockenfels (1997). This model can explain a variety of phenomena associated with laboratory games. In this model, utility functions are of the form $u(x^A, x^A/(x^A+x^B))$. The first argument is known as the absolute payoff, and the second argument is known as the relative payoff. Utility is non-decreasing in the absolute payoff (holding the relative payoff fixed), and strictly decreasing as one moves away from a relative payoff of ½ (holding the absolute payoff fixed). Therefore *A* will want to contribute *1* if he knows *B* will contribute *1*, since this contribution is best for both *A's* absolute and relative payoff. *A* will contribute more than *1* only if *B* contributes

[7] This model mixes linear altruism with the mistakes explanation. While we begin by checking the explanatory power of linear altruism and mistakes separately, later on we consider their joint explanatory power. Another hypothesis that falls into the altruism category is joint payoff maximization. Andreoni and Miller (1996), for example, present evidence that some players are joint payoff maximizers. While we do not explicitly consider this hypothesis in the text, the reader will be able to verify that it cannot explain the shifts in contributions we observe in our experiment.

2. Note that the prediction that *A* chooses the minimum contribution when he knows *B* does so is more restrictive than what Becker's models predicts. We return to this point in the summary section.

Both altruism hypotheses retain C2 and C3. So given linear altruism (C1.a), playing *6*, or playing *1*, depending on the coefficient of altruism *a*, is the best strategy for *A* independent of *B*'s choice. By C2, *A* recognizes this point, and so by C3, *A* plays *6* (or *1*). Linear altruism is strategy insensitive because the predicted contribution is independent of *A*'s expectation of what *B* will do. Distributive preferences, on the other hand, is strategy sensitive because *A*'s contribution depends critically on *B*'s contribution. The experiment allows us to check on this implication.

2.2 Norms of Cooperation: Alternatives to C3

Norms of cooperation assert that people sometimes forgo their own immediate self-interest out of a sense of obligation to conform to particular rules. Hence both norms of cooperation hypotheses reformulate C3. One hypothesis asserts that the right action is independent of what the other player might do (so it is strategy insensitive). The other asserts that the right action depends critically on the action of the other player (so it is strategy sensitive).[8]

> *Unconditional Cooperation (C3.a)*. Player *A* feels obligated to contribute more than the minimum independent of what he expects *B* to contribute.

> *Direct Reciprocity (C3.b)*. Player *A* feels obligated to contribute more than the minimum only if he expects player *B* to contribute more than the minimum.[9]

The key difference between altruism and norms of cooperation is that altruism emphasizes concern for the game outcome, whereas norms emphasize a concern for performing the correct act.

[8]Dawes and Thaler sketch a version of norms of cooperation motivated by altruism. We follow Sugden (1984) because his development offers a sharper alternative to the altruism hypotheses. The reader might wonder how it is possible to achieve self-interested goals by following rules that benefit others. Here is an example: Some religious rules concerning behavior towards others are associated with individual rewards (punishment) for conformity (violation).

[9] The term reciprocity has been used in a variety of different ways, often in connection to multi-period environments. We use the adjective 'direct' to clarify that we refer to a specific definition.

Unconditional cooperation is what Sugden (1984) refers to as 'Kantian' behavior, indicating its philosophical roots. Contributions are conceived as obligations, and as such they are willfully independent of strategic considerations. Distinct from the linear altruism hypothesis, the unconditional cooperation hypothesis allows for a contribution level other than *1* or *6*.

Direct reciprocity is essentially a simplified version of Sugden's (1984) 'principle of reciprocity.' Player *A* feels an obligation to contribute only if *B* also contributes. Likewise, Rabin's (1993) model of fairness equilibrium suggests that *A* will contribute only if *B* does something for *A* (contributes *2*). Sugden (1984) demonstrates that his model is consistent with a set of stylized facts having to do with charitable giving. Rabin (1993) demonstrates that his model is consistent with stylized facts having to do with two-person normal form games. Making contributing *conditional* on *B*'s volition distinguishes direct reciprocity from distributive preferences. *A*'s contribution is a payback for *B* trusting *A* as opposed to a reflection of *A*'s preferences over the payoff allocation. Our experiment exploits this distinction.

2.3 Mistakes: Alternatives to C2

Mistakes hypotheses imply that players cooperate out of confusion or error – a modification of C2. One hypothesis asserts that mistakes are random, while the other asserts they are systematic.

> *Random Errors (C2.a).* Contributing players *A* chooses a contribution randomly drawn from some probability distribution.

This is Ledyard's proposition that "[S]ubjects make mistakes...and choose their allocations randomly." [10] Random errors play a role in several models of public goods experiments, often in combination with some form of altruism (ex., Anderson et al., 1998, and Palfrey and Prisbrey, 1997).

---

[10] We drop the part of Ledyard's statement asserting that subjects "do not care, are bored." The dropped part concerns payoff salience, and implies a modification of C1. Because Ledyard is so specific about the implication

Misidentification (C2.b). Contributing players *A* confuse the game with one that is more common to their everyday experience. Failing to recognize the dominant strategy, they contribute the fixed amount, $g^A > 1$, specified by the social norm they associate with the everyday game.

Misidentification is a specific interpretation of Gale, Binmore and Samuelson's (1995) argument, that they advance in the context of the ultimatum bargaining game,

> Short-term behavior in the Ultimatum Game is likely to be governed by social norms that are triggered by the framing of the laboratory experiment. Rather than being adapted to the Ultimatum Game, such social norms will presumably have evolved for use in everyday cousins of the Ultimatum Game. (p. 72)

Admittedly, we could interpret Gale et al.'s conjecture as implying that the social norm is played intentionally (not a mistake), but this would lead us to the norms of cooperation hypotheses already considered. Misidentification differs from random errors in two ways. First, it postulates a systematic pattern of contribution that random errors does not. Second, misidentification may be strategy sensitive (we elaborate on this below).

Table 1 summarizes the hypotheses we test by key concept and strategy sensitivity.

Table 1. Classification of the Six Hypotheses

	strategy sensitive	*strategy insensitive*
altruism	distributive preferences	linear altruism
norms of cooperation	direct reciprocity	unconditional cooperation
mistakes	misidentification	random errors

of his argument – subjects choose randomly – the C2 formulation is sufficient to capture the empirical content of the argument.

3. The Separation Experiment

The three treatments of the experiment vary the strategy space with the important consequence of varying what *A* knows about *B's* choice of contribution. In this way, the experiment provides a test of the sensitivity of player contributions to other player's contributions. In one treatment, *B*'s choice is unknown, in the second *B*'s choice is known, and in the third *B* has but one choice. We first detail the game in each treatment, and then describe what the hypotheses predict.

The game in the *uninformed treatment* is a standard simultaneous move dilemma game: Player *A* chooses among contributions *1* through *6*, unconditionally and without knowing *B*'s choice. Likewise, *B* chooses between *1* and *2*, unconditionally and without knowing *A*'s choice. Payoffs are as in the Figure 1 matrix.

The game in the *informed treatment* differs from that in the uninformed treatment in but one way: player *A* chooses contingent on *B*'s choice. That is, *A* makes two choices, one for each of *B*'s potential choices. *B* chooses unconditionally, without knowing *A*'s choice. The informed game is strategically equivalent to having *B* choose first, and *A* choose after being informed of *B*'s choice. We use the strategy method; that is, we ask *A* to say what he will choose in both contingencies. This method provides a more complete portrait of *A*'s behavior.[11]

Games in the *dictator treatment* are played precisely as in the uninformed treatment, except both rows of the payoff matrix are identical to the top row in Figure 1; hence *B*'s choice does not matter, making *A* a 'dictator.' The dictator game shares two important features with dilemma games. First, *A*'s dominant strategy is to contribute the minimum possible, *1*. Second, deviations from the dominant strategy increase joint payoffs. The attraction of the dictator game for our purposes is that it is in fact not really a game; it is a one-person decision task. So the

[11]Although standard theory says that the procedures should be equivalent, asking a subject for a conditional choice as opposed to asking after he sees his partner's move may produce differences in results. Schotter, Weigelt and Wilson (1994) provide some evidence on this issue. Brandts and Charness (1998) investigate this issue in experiments with simple sequential games very similar to ours. Their results show no differences between the two elicitation procedures. As shown below, the treatment effects we find are all very large; when there is no effect, this too is unambiguous. It turns out that our results are rather stark and a modest elicitation effect would probably not substantially alter the conclusions we draw.

dictator's choice is free of strategic confounding, and this allows a clear gauge of payoff preferences. The dictator game has been the object of numerous studies (see for example, Forsythe et al., 1994).

Steady framing across treatments is crucial to testing several of the hypotheses we consider. The protocol for the experiment, a complete description of verbal and written instructions provided to subjects, appears in Appendix A. A synopsis is given in section 3.2. For the moment, note that the protocols for the uninformed and dictator treatments are identical save for the necessary difference in the bottom row payoffs in the Figure 1 matrix. The informed treatment differs from the uninformed only in the addition of a few words to explain the conditional choice of *A*.

Subjects play the game once in each player role with no feedback in between games (a detailed description of this design follows in section 3.2). Game repetition was introduced to public goods studies to see if contributions would diminish as players learn about the game and the behavior of other players (Ledyard, 1995, section D). The underlying conjecture is that some contributions are due to confusion about the game or mistaken expectations about the behavior of others. Our experiment allows us to check on these explanations without introducing repetition. Specifically, testing the mistakes hypotheses directly checks on confusion about the game. And we can check on people's expectations about the behavior of others by comparing Player *A* contributions in the uninformed game to those in the other games where *A* conditions his contribution on *B's*.[12]

3.1 Testing the Hypotheses

It is convenient to think of the informed treatment as being composed of two halves, one for *A*'s choice when *B*-chooses-bottom (chooses *1*, the minimum contribution), and one for *A*'s choice

[12]We add that repetition introduces interesting but complicated additional issues, such as strategic reputation building and how people learn. The knowledge gained here concerning why people contribute in the one-time case should be useful to the study of the case with repetition.

when *B*-chooses-top (chooses *2*). So while there are just three treatments, the experiment has four distinct 'cells.'

Table 2. Predictions about Player *A*'s contribution

Hypothesis	Predictions
linear altruism	$g_U^A = g_{I1}^A = g_{I2}^A = g_D^A = 1 \text{ or } 6$
distributive preferences	$g_{I2}^A = g_D^A > g_U^A \geq g_{I1}^A$
unconditional cooperation	$g_U^A = g_{I1}^A = g_{I2}^A = g_D^A$
direct reciprocity	$g_{I2}^A > g_U^A \geq g_{I1}^A = g_D^A$
random errors	$g_U^A = g_{I1}^A = g_{I2}^A = g_D^A$ (expected value)
misidentification	$g_U^A = g_D^A;\ g_{I1}^A = g_{I2}^A$

U = uninformed; I1 = informed, *B*-chooses-bottom row; I2 = informed, *B*-chooses-top row; D = dictator.

Each hypothesis makes a prediction about the pattern of *A*'s contributions as the information about *B*'s contribution changes. In terms of the experiment, each hypothesis makes a prediction about how contributions will shift (or not shift) across cells. Before going on to the explanation the reader should take a moment to examine the summary of predictions about shifts in contributions displayed in Table 2.

We begin with the three hypotheses that are strategy insensitive. All predict that *A*'s contribution will be independent of his expectation of *B*'s contribution. The linear altruism hypothesis predicts $g^A = 1$ or 6 for all cases. The unconditional cooperation hypothesis does not specify a particular contribution level since subjects may differ with respect to their sense of obligation. The random errors hypothesis predicts *A*'s contribution will be drawn from the same distribution, $d(g^A)$, across cells; hence, in terms of expected value, the contribution is predicted to be the same in all cells.

Application of the unconditional cooperation hypothesis to the dictator game is open to the objection that the distinct payoff matrix for the dictator game may call up a different (unconditional) norm of behavior than the other games. If we accept this argument, we can confine our comparison to the two informed cells – surely a fair test of the unconditionality. Likewise, one might argue that framing is important to the (otherwise) random errors we see. But the protocol for uninformed and dictator treatments was identical save for the changes in payoff numbers (Appendix A). The two informed cells necessarily had identical instructions. So again, if we wish, we can confine ourselves to these comparisons. (We will not consider these variations further in the text, but the reader can verify that they are contradicted by the data.)

Each of the three remaining hypotheses is strategy sensitive and each predicts a distinct pattern of contribution shifts across cells.

Distributive preferences implies that contributions in the dictator cell should be identical to those in the *B*-chooses-top cell, since in both cases *A* chooses from precisely the same set of allocations. We explained in section 2.1 why there should be more contributions in the *B*-chooses-top cell than in the *B*-chooses-bottom cell. For the uninformed cell, *A* cannot (plausibly) be certain about what action *B* will take. *A* should never contribute more than he would in the *B*-chooses-top cell, since this is the most he would give no matter what *B* contributes. By the same reasoning, *A* should never contribute less than he would in the *B*-chooses-bottom cell. If *A* contributors are risk averse, which is the usual assumption, then contributions should be strictly lower in the uninformed cell than in the *B*-chooses-top cell – a risk averse *A* contributor will respond to the uncertainty about *B*'s contribution by giving less than he would if he were certain that *B* would contribute. If the assessment of the probability that *B* will contribute is low enough, then contributions in the uninformed cell may be as low as in the *B*-chooses-bottom cell.

Direct reciprocity makes two predictions that differ sharply with distributive preferences: First, direct reciprocity implies that *A* contributes strictly more in the *B*-chooses-top cell than in the dictator cell, where he contributes the minimum. The reason is that direct reciprocity posits that *A* feels an obligation to contribute more than the minimum only if *B* chooses to do so, and *B* has a choice in the informed game, but no meaningful decision in the dictator game. Second, because there is nothing for *A* to reward in the *B*-chooses-bottom cell, *A* contributions should be the same minimum amount as in the dictator cell. A third prediction concerning the uninformed cell is similar to that for distributive preferences: Since *A* is uncertain about *B*'s contribution in the uninformed cell, we expect that *A* contributions in the uninformed cell to be less than in the *B*-chooses-top cell, and no less than in *B*-chooses-bottom.

The misidentification hypothesis predicts the same contribution in the dictator cell as in the uninformed cell because the protocols are the same save for the payoff matrix; hence whatever confusion that is engendered by the frame should be the same across the two cells, and so both games should evoke the same social norm and contribution. (One might argue that contributions should be *less* in the dictator cell since the game is in a mild sense simpler – the fact that the two rows of numbers in the payoff matrix are identical in the dictator cell means that the game requires less strategic analysis. So the weaker prediction is that contributions in the dictator cell will be less than or equal to those in the uninformed cell. We leave it to the reader to verify that this variation is contradicted by the data.) The frame for the two informed cells is identical, so these two cells should also evoke the same contribution.

3.2 Sample Sizes, Subject Pool, and Procedures

Each treatment was comprised of two sessions. Each session involved from 14 to 18 subjects (always an even number). There were a total of 32 subjects in the uninformed treatment, 34 in each of the other treatments. All sessions were run at the Universitat Autònoma de Barcelona. Subjects were recruited from undergraduate classes. Participation required appearing at a special place and time, and was restricted to one session. The chance to earn cash was the only incentive offered.

All subjects were seated so that they could not observe others' choices. The lab protocol (Appendix A) includes all verbal and written communication between monitor and subjects save for individual subject questions-and-answers. The same experimenter was the communicating monitor for all sessions (others silently managed the paper flow). Instructions were read aloud. The payoff matrix, shown in Figure 1, was displayed on a blackboard, and each cell was explained. Subjects played two games without knowing which one would be chosen for payment. *A* and *B* roles were alternated between games. Matching was anonymous and randomly selected for each game. After both games were played, a coin flip chose one game for payoff. To preserve anonymity, those with the *B* role for the chosen game were taken to another room and paid separately. Subjects received no fee beyond the earnings for the game. Average subject earnings were 1134 pesetas, equivalent to about $8.72 at the exchange rate that prevailed at the time of the experiment.[13]

The experiment had a "no feedback" design: subjects were *not* informed of the outcome of the first game prior to playing the second. This arrangement has two advantageous features. First, order and learning effects are not an issue; in this sense each game is 'single play.' Second, roles are rotated between games, allowing us to gather player *A* data from each subject. An important point about both of these features is that they are held constant across treatments, and so they cannot explain the shifts in contributions we observe.

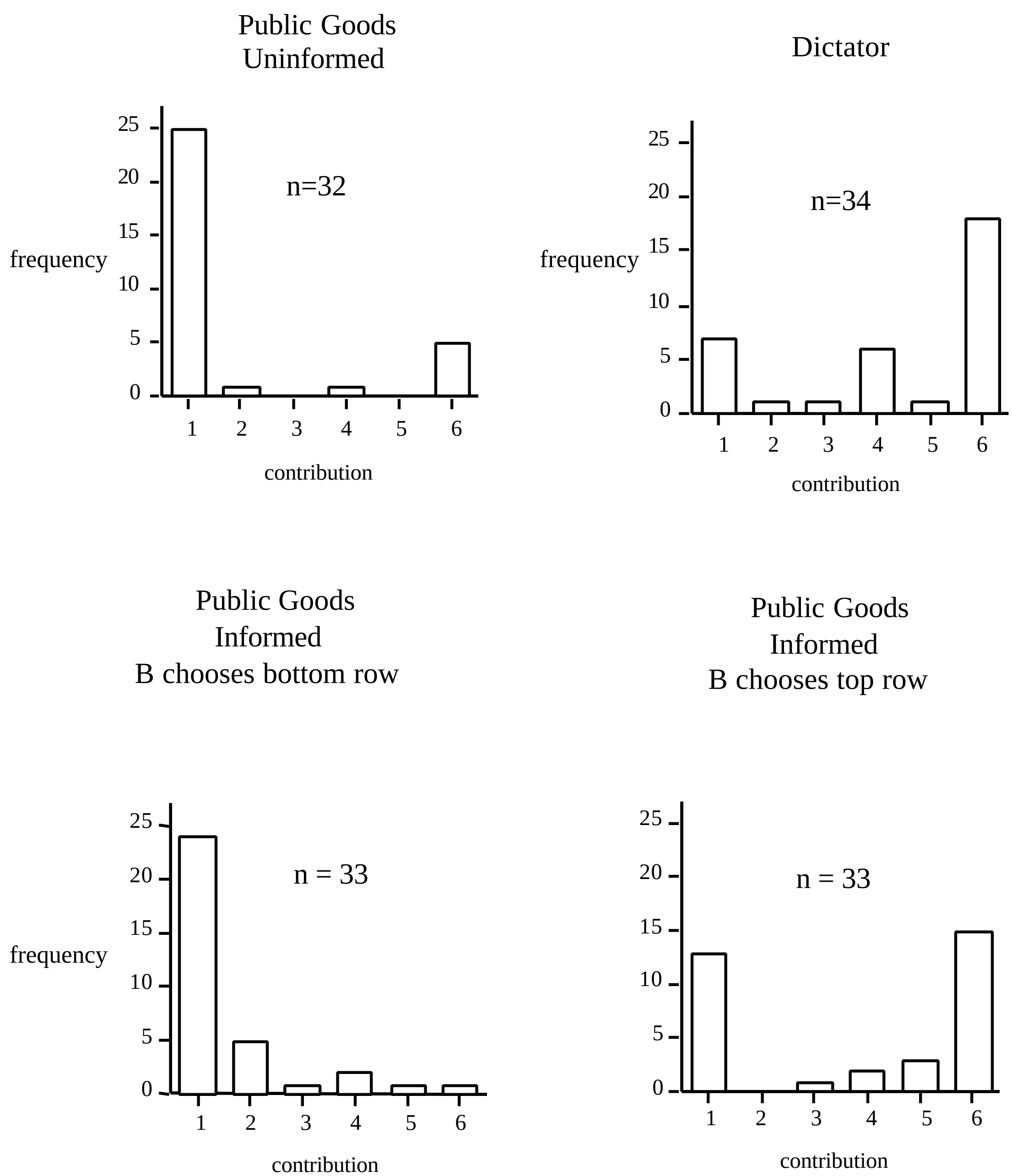
In the context of our game, the *double blind hypothesis*, a methodological hypothesis not considered above, attributes subject contributions to experimenter observation. Laury, Walker and Williams (1995) performed a direct test of double blind in a public goods game, and found no supporting evidence. Most importantly, all treatments were run without a double blind procedure, so the double blind explanation cannot explain the shifts in contributions we observe.

[13] Each session lasted about 45 minutes.

4. Results from the New Experiment

The complete data set appears in Appendix B. Figure 2 displays player *A* contributions by cell.

Figure 2. Player *A* Contributions†



†There are 33, instead of 34, observations for the informed treatment because we eliminated one subject who did not indicate a choice when *B* does not contribute.

Contributions are higher for the dictator cell and the *B*-chooses-top cell; lower in the other two.

Table 3 summarizes the results of a statistical comparison of the distributions.

Table 3. Statistical Comparison of the Four Cells

t-test: difference in average contribution level

$\bar{g}^A$ = sample mean

$\bar{g}^A$(col) - $\bar{g}^A$(row) (one tail *p*-value)	Uninformed	Dictator	Informed: *B*-chooses-top
Dictator	-2.48 (< 0.0001)		
Informed: *B*-chooses-top row	-1.89 (0.0003)	0.59 (0.6357)	
Informed: *B*-chooses-bottom row	0.26 (0.2557)	2.74 (< 0.0001)	2.15 (< 0.0001)

Empirical c^2-test: homogeneity of distribution, $d(g^A)$

c^2-statistic (*p*-value)	Uninformed	Dictator	Informed: *B*-chooses-top
Dictator	23.00 (< 0.0001)		
Informed: *B*-chooses-top row	14.08 (0.0029)	6.07 (0.2961)	
Informed: *B*-chooses-bottom row	7.63 (0.1412)	29.20 (< 0.0001)	21.52 (< 0.0001)

Empirical c^2-test: *p*-values are averages from five 20,000 trial samplings of the contingency table distribution. This method of calculation provides a more accurate *p*-value than a test that relies on a c^2 table approximation.

We investigate the pattern of contributions across cells. A *t*-test of the difference between means checks for a location shift. A chi-square test probes for any general differences. The results are unambiguous, and divide the cells into two distinct sets. Specifically, there is no significant difference between the dictator distribution and the *B*-chooses-top distribution. Likewise, there is no difference between the uninformed distribution and the *B*-chooses-bottom distribution. Contributions are higher in the former two cells than in the latter two (for both tests $p < 0.0003$).

Technically, the two informed distributions are not independent. A better approach to this comparison is a matched pair test: If the two distributions are the same, then we expect the number of subjects with a larger contribution in the cell where *B*-chooses-bottom to be offset by an equal sized group that decrease their contribution. In fact, 18 increase and 2 decrease (the rest stay the same). This is strongly statistically significant (one-tail $p < 0.0001$). So the matched pair test yields a very similar result to that reported in Table 3.

In terms of the Table 2 notation, the comparative static results are

$$g^A_{I2} = g^A_D > g^A_U = g^A_{I1} \tag{4.1}$$

Equation (4.1) is consistent with the distributive preferences hypothesis. The hypothesis accurately predicts the shifts and non-shifts in the contribution distribution in all instances.

The sharp treatment effects leads us to reject all hypotheses that are strategy insensitive: linear altruism, unconditional cooperation and random errors. The strongest evidence against linear altruism is the sharp difference between the two informed cells – the extent of *A*'s cooperation clearly depends on *B*'s contribution. Random errors also fail here, but perhaps the strongest evidence against this hypothesis is the failure to predict that the strategically simpler dictator game will garner a higher contribution level than the uninformed game.

Misidentification is also rejected as an explanation because contributions are much *higher* in the dictator cell than in the uninformed cell, not the same. The difference between informed cells is also inconsistent with misidentification.

Direct reciprocity predicts that contributions should be the same for the *B*-chooses-bottom cell and the dictator cell. This is strongly rejected. Further, the comparison of contributions in the dictator cell and the B-chooses-top cell produces no evidence that the latter is higher, as direct reciprocity predicts. Since this failure is equivalent to failure to reject a null hypothesis, it raises the issue of statistical power. Observe, however, that average contributions were actually somewhat *lower* for *B*-chooses-top (Table 3), the opposite to the direction predicted by the hypothesis. More formally, we can get a sense of the power of the test by formulating direct reciprocity as the null hypothesis, $g^A_{I2} - g^A_D = c$, and then asking, Given the data, what is the lowest value of c for which we can reject? The arithmetic produces $c = 0.35$. So even if we conjecture that the average reciprocal reward for cooperation is just 0.35 of a contribution level (equivalent to a reward of about 70 pesetas, or $0.54), we reject direct reciprocity.[14]

The normal good assumption made in conjunction with the Becker-model version of the distributive preferences implies that contributors in *B*-chooses-bottom would contribute at least two steps higher in either the dictator or *B*-chooses-top cell (inspect the Figure 1 payoffs). Of course some people may not contribute in any cell, and this could drive the average increase in contribution below 2, even if the normal assumption accurately reflects the preferences of the contributors. In spite of this, Table 3 indicates that in comparing *B*-chooses-bottom with either the dictator or the *B*-chooses-top cell, the average increase exceeds 2, unambiguously consistent with the normal good assumption. Testing the normal good assumption by comparing uninformed contributions to dictator or *B*-chooses-top is potentially confounded by the uncertainty surrounding *B*'s contribution in the uninformed cell. Nevertheless, the differences between uninformed and dictator is greater than 2, and the difference between uninformed and *B*-chooses-top is close to 2. In sum, the data is in line with the normal good assumption.[15]

We might ask whether *B* players anticipated the higher *A* contributions when *B*-chooses-top. If so, we would expect that more players would be likely to contribute when in the role of *B*. In fact, 41 percent (14/34) did so. This is weakly significantly different than the 25 percent observed in the uninformed cell (one tail p-value = 0.082).

We now turn to consider whether any additive combination of the reasons for contributing advanced by the five hypotheses that fail individually, can explain the pattern of contributing summarized by equation (4.1). The following statement answers this question, and also summarizes our findings:

---

[14] Bolton, Brandts and Ockenfels (1998) present additional evidence inconsistent with what we have called direct reciprocity.

[15] The aggregate evidence is also consistent with the notion of *cooperative gain seeking* put forward in Brandts and Schram (1996). According to this concept some players are willing to cooperate as long as the payoffs they obtain are higher than would be the case if everybody free rode. The average contribution in the B-chooses-top cell is 3.2 and yields both players a payoff higher than the 750 they would obtain at the standard Nash equilibrium outcome.

Within the set of six hypotheses, distributive preferences is sufficient to explain equation (4.1), and also necessary in that no additive combination of the remaining five hypotheses is consistent with (4.1).

We have already demonstrated the sufficiency part. For necessity, we have already explained that the remaining five hypotheses are individually inconsistent with (4.1). Now note that, excluding distributive preferences, any additive combination of the remaining five hypotheses implies either $g_U^A = g_D^A$ or $g_U^A \geq g_D^A$ (see Table 2), inconsistent with (4.1).

Finally, the reader might wonder whether the distributive preferences hypothesis' success with predicting the shifts and non-shifts in contributing across treatments could be plausibly attributed to serendipity. A simple calculation suggests this is unlikely: The experiment produces four contributor distributions for which there are 6 pair-wise comparisons. We found that the distributions in two of these comparisons to be statistically the same, while the other four comparisons were statistically different (see Table 3). The chances of getting all six comparisons correct by arbitrarily guessing 'same' and 'different' is $½^6$ = .016. Even if we throw out the one instance in which the distributive preferences hypothesis makes a prediction with weak inequality, the chance of getting the experiment right by chance is only $½^5$ = .031.

5. Discussion and Summary

To summarize, we began with six hypotheses for contributions in dilemma games. Each can be classified in two ways; first, by whether the key concept is altruism, norms of cooperation or mistakes; and second, by whether the hypothesis is strategy sensitive, in that contributions are predicted to be dependent on information about others' contributions. The domain for the test was a simple two-person dilemma game with a standard public goods payoff function. We compared the hypotheses on the basis of their ability to track the strategic behavior of contributors; specifically, on the basis of how well a hypothesis tracks changes in the level of contribution in response to changes in information about other player contributions.

The data led us to two important findings. First, the results are conclusive evidence that a great deal of contributor behavior is strategically interdependent. Strategy insensitive hypotheses do not account for the strong shifts in contributions observed when we manipulated information about the other player's contribution. Second, of the strategy sensitive hypotheses, only distributive preferences accurately tracks the pattern of contributions we observe. No additive combination of the other five hypotheses is capable of explaining the pattern.

There are, of course, features of the data within, as well as between, treatments. To a certain extent, the models behind the hypotheses make predictions about these features as well. For example, we described two models to motivate the distributive preferences hypothesis. We noted that the ERC model predicts no one will contribute in the *B*-chooses-bottom cell of the informed treatment, but in fact 27% (9 of 33) do choose to contribute. Becker's model is less specific on this score, and so there is no inconsistency with this model.

Andreoni (1995), Palfrey and Prisbrey (1997) and Anderson, Goeree and Holt (1998) all present analyses of (more than 2-person) experimental public goods games in which mistakes are combined with some kind of purposeful behavior. Anderson, Goeree and Holt (1998) combine linear altruism with mistakes, whereas Palfrey and Prisbrey (1997) explain their data combining heterogenous warm glow altruism and mistakes.[16] These models cannot accommodate the shifts in contributing that we observe across treatments. Andreoni's (1995) 'kindness' interpretation of his experiment is compatible with the pattern of shifts we observe, although Andreoni's interpretation is not specific about the nature of the 'kindness' involved.

This might be a good place to reiterate a point we made in the introduction: Our experiment does not rule out some contributors behaving in accord with any particular explanation. So it may well be that the larger than minimum contributions in the *B*-chooses-bottom cell are due to linear altruism or mistakes (to name two). The important point is that these particular explanations are inadequate to explain the strategic behavior on display in our experiment. A sizable number of

[16] In the context of our experiment, the utility function Palfrey and Prisbrey's identify evidence for is of the form $u(x_A, g) = x_A + ag$ where g is the level of gift (1 through 6 in our experiment), and a is a parameter. These preferences imply that the contributor will be strategy insensitive.

individuals apparently have reasons for contributing that these hypotheses do not account for. Any model that neglects these reasons runs the risk of missing the strong strategic interdependence exhibited among contributions.

As we also pointed out in the introduction, it may be that the strategic nature of contributing is different when we move to a dilemma game with more than two players. That this nature is sharply discontinuous immediately above two, however, strikes us as implausible. We would pose our skepticism in terms of the hypothesis that best fit the strategic behavior we observed here: 'Why would a player care about the distribution of payoffs when there is one other player, but not when there are two or three others?' It strikes us as more plausible that, if the nature of strategic behavior changes with more players, it changes gradually. The important implication being that strategic behavior in the two-player game is related to that in the three or four (or more) player games.[17]

One final implication: At the outset, we noted that there was another sense in which one might ask for an explanation for contributions. Specifically, one might ask whether contributions are, perhaps in some disguised way, consistent with strictly self-interested behavior. While we did not deal specifically with this first-principles question, one of our major findings provides theorists who wish to dig deeper with a clue as to what a good first principle theory will have to look like. The particular pattern of contributing that we observe indicates a fair amount of generosity, but also indicates that contributors are highly averse to being strategically exploited. In this sense, the contributor behavior we observe has a strong self-interested component.

[17] An apparent question that the conjectured continuity raises is what a player would condition their own contribution on when there is more than one other player's contribution to account for. One possibility that strikes us is some summary statistic of other's contributions, such as the average.

References

Anderson, Simon P., Jakob K. Goeree and Charles A. Holt, "A Theoretical Analysis of Altruism and Decision Errors in Public Goods Games," *Journal of Public Economics*, 70, 1998, 297-323.

Andreoni, James, "Cooperation in Public-goods Experiments: Kindness or Confusion?" *American Economic Review*, 89, 1995, 891-904.

Andreoni, James and John H. Miller, "Giving According to GARP: An Experimental Study of Rationality and Altruism," working paper, 1996.

Becker, Gary S., "A Theory of Social Interactions," *Journal of Political Economy*, 1974, 82, 1063-1093.

Bolton, Gary E, Jordi Brandts and Axel Ockenfels, "Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Games," *Experimental Economics*, forthcoming.

Bolton, Gary E and Axel Ockenfels, "ERC: A Theory of Equity, Reciprocity and Competition," working paper, 1997.

Brandts, Jordi and Gary Charness, "Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games," working paper, 1998.

Brandts, Jordi and Arthur Schram, "Cooperative Gains or Noise in Public Goods Experiments," discussion paper Tinbergen Institute 96-81/1, University of Amsterdam, 1996.

Dawes, Robyn M. and Richard H. Thaler, "Cooperation," *Journal of Economic Perspectives*, 2, 1988, 187-97.

Edgeworth, Francis Y., *Mathematical Psychics*, Kegan Paul, London UK, 1881.

Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton, "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior*, 6, 1994, 347-69.

Gale, John, and Kenneth G. Binmore, and Larry Samuelson, "Learning to be Imperfect: The Ultimatum Game," *Games and Economic Behavior*, 8, 1995, 56-90.

Hochman, Harold M. and Rodgers, James D., "Pareto Optimal Redistribution," *American Economic Review*, 59, 1969, 542-57.

Isaac, R. Mark and James M. Walker, "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism," *Quarterly Journal of Economics*, 103, 1988, 179-199.

Laury, Susan K., James M. Walker and Arlington W. Williams, "Anonymity and the Voluntary Provision of Public Goods," *Journal of Economic Behavior and Organization*, 27, 1995, 365-80.

Ledyard, John, "Public Goods: A Survey of Experimental Research," in *Handbook of Experimental Economics* (John H. Kagel and Alvin E. Roth, eds.), Princeton: Princeton University Press, 1995, 111-94.

Offerman, Theo, Josep Sonnemans, and Arthur Schram, "Value Orientations, Expectations, and Voluntary Contributions in Public Goods," *The Economic Journal*, 106, 1996, 817-845.

Palfrey, Thomas R. and Jeffrey E. Prisbrey (1997), "Anomalous Behavior in Public Goods Experiments: How Much and Why?" *American Economic Review*, 87, 829-846.

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

Schotter, A., Weigelt, K., and Wilson, C. (1994). "A Laboratory Investigation of Multiperson Rationality and Presentation Effects." *Games and Economic Behavior*. 6, 445-468.

Sugden, Robert, "Reciprocity: The Supply of Public Goods through Voluntary Contributions," *The Economic Journal*, 94, 1984, 772-87.

Appendix A: Laboratory Protocol

This section contains a description of all procedures, as well as verbal and written instructions given to subjects for the uninformed treatment. The monitor read all verbal instructions directly from the protocol. The only monitor-subject communication not included in the protocol are answers to individual subject questions. The dictator treatment differed solely in the numbers in the bottom row of the Earnings Table. The informed treatment differed only as indicated on written instructions and task forms. The protocol has been translated from the original Spanish.

Seating. Upon entering the room each participant is randomly directed to a seat. Half the seats have red folders and half have blue. Red folder seats on the right side of the room, and blue folder seats on the left side. May I have your attention please. We are ready to begin. Thank you for coming. With the exception of the folder, please remove all materials from your desk. Open your folder and take out the sheet marked 'Instructions'. At this time please read the instructions. *Participants read silently. Written instructions begin here:*

Instructions

General. The purpose of this session is to study how people make decisions in a particular situation. Feel free to ask a monitor questions as they arise. From now until the end of the session, unauthorized communication of any nature with other participants is prohibited.

During the session you will make money. Upon completion of the session the amount you make will be paid to you in cash. Payments are confidential: no other participant will be told the amount of money you make.

Half the participants were given blue folders and half red folders. During the session, you will be paired with a person having a different color folder than yours. No one, however, will know the identity of the person they are paired with. Nor will these identities be revealed after the session is complete.

Decision Task. In each pair, one person will have the role of *A*, and the other will have the role of *B*. The amount of money you earn depends on the decision you make and the decision of the person you are paired with. The Earnings Table below describes the options available to each person and the associated earnings. You make your decision by choosing one of the options available to you and recording it on a paper form. The person in role *A* can choose from options A1 through A6. The person in role *B* can choose option B1 or B2. Each person makes their decision without knowing the decision of the other. [*In the informed treatment only:* B decides unconditionally; that is, he simply chooses between B1 and B2. A chooses conditionally; that is, A indicates a decision for the case where B has chosen B1 and a decision for the case where B has chosen B2. The decision of A that counts is the one that corresponds to the decision of B. Each person receives the earnings in the Earnings Table cell corresponding to the chosen options.

(Same earnings table as in Figure 1 appeared here.)

Conduct of the Session. You will participate in two decision tasks, called Task 1 and Task 2. Both tasks are identical to the description in the previous paragraph. For each task, you will be paired with a different person.

You will have the role of *A* in one task, and the role of *B* in the other. In task 1, those with red folders will have the role of *A*, and those with blue folders will have the role of *B*. In task 2, blue folders are *A*, and red folders are *B*. First, you will receive a decision form for the role you have in task 1. You will complete task 1, and all the forms will be collected. You will then receive a decision form for task 2 and complete task 2. The results for task 1 will not be revealed prior to completion of task 2.

Payment. You will actually be paid your earnings for just one of the two tasks. The one for payment will be chosen by a coin flip after both tasks have been completed. In order to preserve anonymity, the participants who are *B* for the chosen task will be taken to a second room and paid separately. Once you are paid, you may leave. *Written instructions end here.*

I will now read the instructions out loud. *Read instructions*. Are there any questions?

Performing the decision task. Pass out decision forms for task 1. We are ready to begin decision task 1. You make your decision by filling out the form that is now being handed out. Those with red folders will receive the form for role A, and those with blue folders will receive the form for role B. Please review the form with me. *Written decision form for A begins here (form for B was analogous).*

Distribution Task 1

You have the role of *A*.

The person you have been paired with has the role of *B*. You must make your decision by choosing one of the options available to you and described in the earnings table below.

(Same earnings table as in Figure 1 appeared here.)

Use the information in the earnings table to make your decision by circling the number of your chosen option.

A's Decision: Option A1 A2 A3 A4 A5 A6

[*In the informed treatment substitute: A's* Decision if *B* chooses B1: Option A1 A2 A3 A4 A5 A6
A's Decision if *B* chooses B2: Option A1 A2 A3 A4 A5 A6]

When finished, please turn this sheet over and wait quietly. *Written decision form ends here.*

Are there any questions?...Please fill out the form. Again, when you are finished, turn the form over and wait quietly. *Wait until all are finished.* A monitor will now come around to collect the forms. *Collect the forms.* We are now ready to begin decision task 2. Remember, you will be paired with a different person than you were in task 1. *Hand out forms for task 2.* Are there any questions?...Please fill out the form. Again, when you are finished, turn the form over and wait quietly. *Wait until all are finished.* A monitor will now come around to collect the forms. Collect the forms *A coin is flipped to determine roles. Those in role B are taken to a different room and paid separately after those in role A.*

Appendix B: The Data

The following three tables present all the data for the experiment. Each row of each table displays responses for one subject, for both *A* and *B* roles. *A-yes* refers to *A*'s response when *B* chooses the top row; *A-no* the response when *B* chooses the bottom row.†

Uninformed

	A	*B*
	1	1
	1	1
	1	1
	6	2
	1	1
	1	2
	1	1
	1	1
	1	1
	1	1
	1	1
	1	1
	1	1
	1	1
	1	1
	2	1
	1	2
	1	1
	6	1
	1	1
	6	2
	1	2
	1	1
	1	1
	1	2
	1	1
	6	2
	1	1
	4	1
	6	1
	1	1
	1	2
mean	1.9	1.3
std dev	1.9	0.4

Dictator

	A	*B*
	4	1
	1	2
	6	1
	5	1
	6	2
	6	2
	6	2
	6	2
	6	1
	6	1
	6	1
	1	2
	4	1
	3	2
	6	1
	4	1
	1	2
	4	1
	4	2
	6	2
	6	1
	6	2
	4	2
	1	2
	6	2
	6	2
	6	2
	2	1
	1	2
	1	1
	6	1
	6	2
	1	1
	6	2
mean	4.4	1.6
std dev	2.0	0.5

Informed

	A-yes	*A-no*	*B*
	6	1	1
	1	1	1
	1	1	1
	6	2	2
	1	1	1
	1	1	1
	6	1	2
	4	4	2
	6	1	1
	4	1	2
	6	2	2
	1	1	1
	1	1	2
	6	1	1
	6	1	1
	6	4	2
	1	1	2
	5	2	2
	1	1	1
	1	2	1
	1	1	1
	6	1	2
	6	1	1
	1	1	1
	6	1	1
	5	2	2
	6	1	1
	1	1	1
	3	*	1
	6	1	1
	6	5	2
	1	1	1
	5	6	2
	6	1	2
mean	3.8	1.6	1.4
std dev	2.3	1.3	0.5

*This value left blank by the subject. Data for this subject was omitted from the analysis reported.

†An earlier version of this appendix listed *B* moves of 1 as 2, and 2 as 1. *A* moves were correct as originally listed, and all data analysis in the text was correct as originally presented.