DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

Anomalous Behavior in Linear Public Goods Experiments: How Much and Why?

Thomas R. Palfrey
California Institute of Technology

Jeffrey E. Prisbrey
Universitat Pompeu Fabra, Barcelona, SPAIN

SOCIAL SCIENCE WORKING PAPER 833

January 1993

Anomalous Behavior in Linear Public Goods Experiments: How Much and Why?

Thomas R. Palfrey Jeffrey E. Prisbrey

Abstract

We report the results of voluntary contributions experiments where subjects are randomly assigned constant marginal rates of substitution between the public and the private good. These random assignments are changed after each decision period. The design allows us to measure the response functions of the players in much the same way that bidding functions can be measured in private good, sealed-bid auction experiments. The results are quite different from the results of others in environments with little or no heterogeneity. We see much more free riding, very little evidence of decay across periods, and only sparse evidence of anomalous behavior such as splitting, spite, and decay.

Keywords: Voluntary contributions, public goods, experiments

JEL Classification numbers: 026, 215

Anomalous Behavior in Linear Public Goods Experiments: How Much and Why?*

Thomas R. Palfrey   Jeffrey E. Prisbrey

1 Introduction

There is a growing body of data obtained from experiments on voluntary contributions in linear public goods environments with a single public good and a single private good. Many features of the data have been difficult to explain; for example, subjects violate dominant strategies on a regular basis. They give away money, apparently just to be nice (Isaac and Walker [1984, and elsewhere]); at least as often, they seem to give away money just to be mean (Saijo and Yamaguchi [1992]). Furthermore, individual behavior over time exhibits erratic patterns, it alternates back and forth between extreme generosity and extreme selfishness. Ledyard's (1992) excellent survey documents these and several other anomalies.

These anomalies might be cause for alarm as they signal trouble for any but the most schizophrenic models of behavior. However, the range of enviornments for which these experimental results have been reported is very narrow, and the designs employed make it difficult if not impossible to identify decision rules at an individual level. The point of this paper is to broaden the playing field in a natural direction, using a design that permits estimation of individual behavior. By changing both the information structure and the distribution of preferences, this design also provides a robustness check on the anomolous findings of past experiments.

We offer the following thought experiment in the context of a well-studied private goods allocation mechanism, the second-price auction, in hopes that it will help the reader understand some of our concerns about design, and to foreshadow what follows.

*The financial support of the National Science Foundation is gratefully acknowledged. We thank John Ledyard for his comments on an earlier draft, and Mark Isaac and Jimmy Walker for sharing their data and for offering many helpful suggestions and comments.

A Thought Experiment:

> Imagine conducting a second-price sealed bid auction experiment with four players, where each is told to bid for an object that is worth exactly \$1.58 to him. After careful explanation of the rules, ten identical, sealed bid, second-price auctions are then conducted in sequence. Bids are required to be greater than or equal to 0 and less than \$1.58 and ties are broken randomly. After each auction, subjects are told the winning bid and the second highest bid. When the tenth auction is over, everyone is paid by the experimenter and thanked for showing up.
>
> What do you think the distribution of bids will be, and how will this distribution change from period to period? How would you plan to bid in such an auction?

The first observation to be made about the thought experiment is that it shares some of the traits of many voluntary contribution, public goods experiments that have been reported in the literature. In the most common voluntary contribution, public goods experiment, like in the thought experiment, there are a number of identical players. Also, the players are asked to make a decision about buying a good and they are given personal incentives not to buy it, or at least to spend as little as possible on it. Much of what is known about free riding is based on experiments with this type of design.

The second observation to be made is that little can be learned about the general bidding behavior of the participants. In the auction, each player attaches the same value to the good in each of the ten auctions. Furthermore, every other player also attaches this same value to the good. The measurement of a general bidding function is practically impossible; the best one can do is estimate behavior at a particular point.

It would be possible, by running a number of experiments and varying the value of the good, to construct something that looked like a bidding function. However, that function would depend upon the fact that every player attaches the same value to the auctioned good. This function would only measure how an individual's choice behavior changes when their own value *and the joint distribution of all bidders' values* change simultaneously. The estimated function would have other limitations as well—to obtain the data required, an individual would have to participate in a large number of 10-auction sequences. The amount of play necessary might lead to a confounding of the effects of bidding behavior and of experience, unless a large number of experiments were conducted.

A final observation is that, in spite of the fact that there is a dominant strategy equilibrium where each bids \$1.57, one can, for a variety of reasons, imagine players bidding differently. In fact, it is difficult to guess what might actually happen, especially if the players are inexperienced.[1]

[1] One might also notice that the thought experiment is a repeated game not a one-shot game. We do not address this potential complication until later in the paper.

It should be no surprise to learn that auction experiments are not usually conducted like the thought experiment. Auction experiments have focused exclusively on different environments, environments in which players have diverse preferences and diverse information. These are the environments in which auctions most naturally occur. What is surprising is that voluntary contribution experiments have, for the most part, not shared this focus.[2]

This paper, and the experimental design it employs, is motivated by our reflections about the thought experiment, and by a view that much can be gained by shifting the research agenda in the direction of this different class of environments. One benefit is simply better measurement: response (*bidding*) functions can be estimated at the individual level. Also, we can check for the robustness of existing results to environments that include features, such as heterogeneity of preferences, that are endemic to natural settings. In what follows, we report results from our experiments that study this kind of environment, and we contrast these results with previous findings.

2 Background

This paper investigates contribution behavior under the Voluntary Contribution Mechanism in simple linear public good environments where all players have dominant strategies. The typical environment consists of N individuals, each endowed with X_i discrete units of a private good. The marginal rate of transformation between the public good, y, and the private good is one-for-one, and individual utility functions are of the form: $U(y, x_i) = Vy + r_i x_i$. We refer to V as the *value of the public good*, and it is normalized to be the same for all individuals.

The Voluntary Contribution Mechanism defines a simple game, in which each individual simultaneously decides how much public good (between 0 and X_i) to produce on his own. Total public good production in the economy is the sum of all private production of the public good. Payoff functions are then defined from the final allocation and the utility functions in the obvious way. This game is repeated several times.

As pointed out in Section 1, much of what we think we know about behavior in this game is based on experiments in which X_i and r_i are the same across individuals and repetitions and $r_i/V > 1$. This paper concentrates on a group size of four.

[2]There are a few exceptions, notably Fisher et al. 1991 and Isaac et al. 1985, both of which consider environments with two types. The former provides subjects with identical information about other subjects' preferences as in parallel homogeneous preference experiments. The latter has several other different features, including nonlinearities, and does not conduct any baseline experiments with homogeneous preference. Brookshire et al. (1991), Smith (1980), and Marwell and Ames (1980) also have conducted experiments with heterogeneous preferences, but these are not comparable for other reasons. None of these experiments varied individual subject preferences across decisions, nor did they provide explicit information about the distribution of preferences in the population. Palfrey and Rosenthal (1991) use an environment similar to the one explaned here, but the public good technology is step-level, not linear.

Several findings have emerged from these other investigations: (1) nearly all players in this game violate their one-shot dominant strategy, with many contributing upwards of half their endowment, even when r_i/V is three or more; (2) there is a strong negative relationship between the marginal rate of substitution r_i/V and the rate at which violations are observed; (3) roughly half the aggregate private endowment is contributed by inexperienced subjects on the first play of the game; (4) violations of dominant strategies diminish with repetition and with experience (playing a second sequence of games with a new group); (5) violations of dominant strategies to contribute ($r_i/V_i < 1$, Saijo and Yamaguchi [1992]) appear to be even more prevalent than violations of dominant strategies to free ride.

3 Our Design and Procedures

Our experiment looks at the above findings more closely by studying evironments with both non-degenerate distributions of r_i/V, and with private information. These innovations are introduced to overcome the limititations of past designs, limitations suggested by the thought experiment. The innovations permit us to measure responsiveness to r_i/V, via response or bidding functions, at both the individual level and the aggregate level, and to measure a baseline of deviant or erroneous behavior due to nuisance factors, such as boredom or confusion.

There are a number of specific features of our design that enable us to address other issues that are relevant to understanding other commonly observed patterns of behavior. These features are listed below. A sample copy of the instructions is in the Appendix.

1. In all our environments, subjects receive r_i's that are randomly assigned according to a uniform distribution between 1 and 20. We sometimes refer to these as *token values*. Each time a subject is to make a new decision, he is independently and randomly assigned a new r_i for that decision. Subjects do not know the other subjects' assignments of r_j's, but the distribution is publicly announced at the beginning. The value of V is also announced at the beginning.

   Therefore, the data contain multiple observations of the choice behavior of each individual, observations at different levels of r_i/V, and permits the estimation of response functions at both the individual and aggregate levels.

2. We vary the distribution of marginal rates of substitution, (r_i/V), by shifting V. We look at the four different distributions given by $V \in \{3, 6, 10, 15\}$. One of the distributions, $V = 3$, has the feature that group efficiency is *not* maximized when all subjects contribute in every round. In that condition, on average, forty percent of the time subjects are assigned a token value that is worth more than four times the individual marginal value of the public good. In these cases, contribution *reduces* group efficiency.

3. We vary the endowment. In one condition, everyone is endowed with one indivisible unit of the private good. In the other condition everyone is endowed with nine discrete units.

4. Each subject makes a sequence of ten decisions in a fixed group with three other players. This allows a direct comparison to some past experiments, notably those reported in the Isaac and Walker studies.

5. Each subject participates in a total of four sequences, each time with a different group of subjects. The first two sequences have the same parameters; the last two sequences have the same parameters (but different from the first two). This allows us to identify experience effects. All four sequences occur in a single session that lasts approximately $1\frac{1}{2}$ hours. Each session includes sixteen subjects.

6. All sessions were conducted at the Caltech Laboratory for Experimental Economics and Political Science, using a collection of PC's that are linked together in a network.

7. Each subject was paid cash, based on a session-specific exchange rate, for each point they earned in the session. The exchange rate was picked so that the sum of equilibrium payoffs was approximately the same across sessions.

[Table 1 here]

4 Response Functions and Background Noise

We focus mainly on two aspects of the data. The first has to do with attempting to identify what we call *errors* or background noise—behavior that is grossly inconsistent with standard theory. Second, we attempt to measure response functions, which are the analog to bidding functions in auctions. The functions answer the question: How do contribution decisions depend on the marginal rate of substitution? We measure errors and response functions at both the aggregate and individual levels, using nonparametric and parametric models of the error structure.

It is useful to think of our analysis in the context of a random utility model, of the sort found in Maddala (1983), McFadden (1982), and elswhere, for the analysis of data with limited dependent variables. For example, in the condition where subjects have a single indivisible unit of the private good, they face a simple binary decision. We model the statistical structure of residuals by assuming that utility functions have a random component that is not observed. For lack of a better name, we call this the *altruism* (or *warm glow*) term. Depending upon the value of the altruism term, subjects may receive some additional utility from contributing a unit of their endowment, over and above the utility induced by the payment method used in the experiment.

Theoretically, an optimal response function for an individual with an additive warm glow term, ε_i, is to contribute X_i if $r_i/V < 1 + \varepsilon_i$, and to contribute 0 if $r_i/V > 1 + \varepsilon_i$. Any behavior is optimal when $r_i/V = 1 + \varepsilon_i$. This is what we call a *cutpoint strategy* (Palfrey and Rosenthal [1988]). In fact, this optimal strategy is a one-shot dominant strategy for any values of ε_i, r_i, V, and X_i.

If the value of ε_i is stochastic, and varies according to some assumed distribution, an estimated response function gives the probability of contribution as a function of other controlled variables, such as experience, etc. In addition, the response function gives us indirectly an estimate of "background noise." We look at the effect of the following variables on response functions:

- The induced marginal rate of substitution (r_i/V).
- Experience.
- Endowment (divisible or indivisible – i.e. one or nine units).
- The value of the public good (V).
- Repetition (Is there a decay over the ten rounds of play?).

5 Analysis of the data

5.1 Some baselines

We present three different baseline error rates. This gives a rough calibration of a lower bound on the amount of *background noise*[3] in the experiment. By this, we mean the percent of observed decisions that appear incongruous with nearly any currently accepted theory of rational decisionmaking. We also make an attempt to compare our baseline with baselines observed elsewhere, to the extent possible.

5.1.1 Splitting

By splitting, we mean that a subject contributes some fraction of his endowment, but not all of it. Because of the linear structure of the environment, such behavior is not rational even if a subject has a warm glow term added to his marginal rate of substitution. While it might be possible to think up models where such behavior is rational, such explanations would likely be quite contrived. Tables 2, 3, and 4 present the splitting data from our experiments. Recall that in half of our experiments, subjects were not capable of splitting, since they had only a binary choice. Thus, the data in this table is based on only half the sample. One can see two striking features. First, splitting is more prominent among

[3]Contemporaneous work by Andreoni (1992) is also pursuing this issue.

inexperienced subjects and in the early periods of each 10-period game. Second, splitting almost never occurs when subjects have $r_i/V < 1$. In other words, almost all splitting can be accounted for by subjects who have a dominant strategy to free ride.

[Table 2, Table 3, and Table 4 here]

These findings contrast sharply with those of Issac and Walker. They observe splitting well over half of the time in their data and, for their marginal rate of substitution, or MRS, of 1.33 experiments, there is very little decay of splitting over the course of the ten periods.

[Table 5 here]

5.1.2 Spiteful behavior

Many have speculated that subjects violate their dominant strategy to free ride because of some form of altruism, or alternatively, because their utility function depends on group payoffs in a positive way. If this is the main driving force behind the past findings, then we should see very little free riding when subjects have $r_i/V < 1$. Based on this scenario, violations of dominant strategies to contribute can reasonably be attributed to effectively random behavior. This gives us a second kind of baseline error rate. In our experiments, four percent of the decisions violate the dominant strategy to contribute when $r_i/V < 1$. This number is remarkably stable across periods and across the experience treatment (see Table 6).

[Table 6 here]

5.1.3 Sacrificial behavior

In one of our designs, $V = 3$, the group optimum is not obtained by everyone contributing for every possible r_i they might draw. In particular, the group payoff is maximized if subjects contribute if and only if $r_i \leq 4V = 12$. A subject who contributes when $r_i > 12$ sacrifices more than the entire group benefits. It is hard to imagine any except the most fervent altruists contributing under these circumstances. The frequency of this type of contribution also provides, in a slightly different way, a lower bound on the amount of "crazy" or random behavior. As Table 7 shows, this kind of behavior is approximately as common as spiteful behavior, but virtually disappears with experience (1 observation out of 129).

[Table 7 here]

5.2 Estimation of response functions from aggregate data

5.2.1 A Simple Model

We measure response functions as the probability of contribution as a function of the marginal rate of substitution or MRS. First, consider the following family of theories, a family that includes both the dominant strategy (game) theory and the altruism theories based on an additive warm glow altruism term. Each member of this family is characterized by an error rate, ε, and a threshold, M. An (ε, M) theory states that "Individuals contribute to a public good if and only if the marginal rate of substitution (token value divided by public good value plus warm glow) is less than or equal to M. However, they make errors at a rate of ε."

If $M = 1$, then this is just the dominant strategy theory, modified appropriately to account for the possibility of error. If $M > 1$ this indicates some degree of altruism, everyone is altruistic. If $M < 1$, this indicates negative altruism. According to our data, what is the best theory in this family? Using the criterion of maximum likelihood, the answer is the M^* that produces the fewest classification errors in the data, together with ε^* equal to whatever the classification error generated by M^* is. This is not only easy to calculate, it is also easy to illustrate graphically. Figure 1 displays the answer: In our data, the best theory is $M = 1.1$. It results in only 12.5 percent ($\varepsilon*$) classification errors and is very close to the selfish cutpoint equal to 1.0. Figures 2 and 3 break this analysis down across the various levels of the V-treatment and the two levels of the endowment treatment.

5.2.2 Probit Analysis

An alternative, more familiar way to estimate response functions is by Probit analysis. In effect, the Probit analysis fits curves through the raw data shown in Figures 4-7. In this analysis, we assume that an altruism term, ε_{it}, is a Normally distributed random term added to an individual's MRS that it is independently distributed across individuals and across decisions.

The impact of experience, endowment and other experimental treatments are easily assessed by introducing dummy variables. The simplest probit model, with only a constant term and r_i/V, or MRS, entering on the right hand side gives us an estimate of the average altruism term, which we denote by $\bar{\varepsilon}$, and its standard deviation $\sigma_{\bar{\varepsilon}}$.

We consider five Probit Models which are built by recursively adding independent variables to the basic model. Note that an observation in these models is a decision involving a single token. In order to maintain equal representation between the conditions with an endowment of one and those with an endowment of nine, an investment decision in the endowment of one conditions is given the same weight as nine similar investment decisions in the endowment of nine conditions.

The intercept coefficients in a Probit model represent changes in $\bar{\varepsilon}/\sigma_{\mathcal{E}}$ and the slope coefficients represent changes in $-1/\sigma_{\mathcal{E}}$. The estimated mean, $\bar{\varepsilon}$, is equal to minus the slope coefficient divided by the intercept coefficient. It follows that a negative change in the already negative slope coefficient leads to a decrease in $\bar{\varepsilon}$, holding everything else constant. This decrease is implied by the decrease in variance due to the more negative slope coefficient. If everything is to stay the same, $\bar{\varepsilon}$ must also decrease. The decrease in variance also makes the slope of the curve steeper.

From each Probit Model, we can obtain a response function $\mathcal{P}(\cdot)$, which returns the probability that a subject invests in the public good. The six variables in the other models are: *exper.s*, a slope dummy for subjects with experience; *exper*, a constant dummy for subjects with experience; *endow.s*, a slope dummy for treatments with an endowment of nine; *endow*, a constant dummy for subjects with an endowment of nine; V, the marginal return from the public good; and *period* which ranges from 1 to 10. Coefficients, t-statistics, log likelihoods, and the percentages correctly predicted for each model are given in Table 8.

[Table 8 here]

Turning to specific models, even the simple model $\mathcal{P}_1$, in which a player's investment decision depends only upon MRS, is able to correctly predict 83.064 percent of the observations.

In model $\mathcal{P}_2$, the slope coefficient for the experience variable, *exper.s*, is negative which means the response curve for experienced subjects is steeper than the response curve for inexperienced subjects. The coefficient for the intercept variable for experience, *exper*, is positive. This tends to offset the change in $\bar{\varepsilon}$ implied by the reduced variance, however, the total change in $\bar{\varepsilon}$ is still negative.

A player's cutpoint is the point at which he is indifferent between investing in the public good and investing in the private good, the point where $\mathcal{P}_i = 1/2$. For inexperienced subjects, the estimated cutpoint is 1.641, and for experienced subjects, it is 1.399. This finding reinforces the findings of Isaac and Walker. Experienced subjects are more consistent with the dominant strategy model than inexperienced subjects. In this case, the effect is even significant. Of independent interest is that experienced subjects' response functions are *steeper*, indicating less random behavior.

Probit model $\mathcal{P}_3$, shows a minor effect of the addition of a pair of endowment variables, both equal 1 if the endowment is nine tokens and 0 if the endowment is one token. In this case, the slope shift is positive and the intercept shift is negative. The consequence is that the response function for subjects in the high endowment condition is flatter than the response function for subjects in the low endowment condition. The negative intercept is enough to counteract the higher variance, however, and the high endowment means are less than the low endowment means. The magnitudes of these coefficients are much smaller than those associated with the experience effect and the effect of the endowment change is similarly smaller.[4] The actual differences are shown in Figure 8.

The variable V, which is added in model $\mathcal{P}_4$, measures the marginal valuation of the public good. One interpretation (since we have controlled for MRS) is that its coefficient tells us what happens to a subject's behavior as the payoffs rise. Although the effect is very small, we find that a player's response function becomes steeper, and the average deviation becomes smaller. A similarly small result holds when the period of the decision is taken into account. Holding everything else constant, a player is less likely to contribute in later periods than in earlier periods.

Quite clearly, the major effects are due to MRS and experience. While the endowment condition has some effect, it is not as important. The effects due to the size of the payoffs and to the period of the decision pale in comparison.

5.3 Response Functions and Errors: Individual Level Analysis

The analysis in the previous section assumes that individuals are identical. In fact, there are indications of heterogeneity in our data. Similar indications have also been noted in past work. This section offers a simple approach to look at differences between individuals, based on minimization of classification errors (as in section 5.2.1). We do two things. First, we break down that analysis by individual, and obtain a distribution of classification minimizing cutpoints for individuals. This allows us to identify the fraction of subjects who behave consistently with the Nash equilibrium, subjects we call *Nash players.* Second, from these estimated individual cutpoints, we can obtain a distribution of the error rates across individuals. This gives us a way to identify what fraction of subjects are behaving consistently with *some* cutpoint model.

We define a *Nash Player* as a player who is rational and non-altruistic.[5] That is $\bar{\varepsilon}_i = 0$. With this in mind, consider Tables 7 and 8 which report, by subject, the raw number of classification errors for each of the twenty possible cutpoints. These cutpoints correspond to the possible token values. They are the only applicable cutpoints, because they relate directly to every possible realization of r_i.

[Tables 7 and 8 here]

Each possible cutpoint is given a score based on how well it represents that subject's decisions in the experiment. The score is simply the number of times a violation would

[4] The magnitudes are comparable because the variables, both dummies, are of the same scale, namely 0 or 1.

[5] Because our estimation allows for errors, a Nash Player may be different than a player who *perfectly* follows the decision rule implied by the self-interested model. The difference is that a Nash Player is allowed to make mistakes.

have occurred if that was the actual cutpoint rule the subject used.[6] More specifically, we hypothesize that a particular player is using a cupoint that corresponds to token value x (we consider every possible x in turn). Hypothetically, each time that player receives a token value r_i, he compares it to x and then spends only if $r_i < x$. A classification error occurs if one of the two following events occurs: $r_i < x$ and the player does not spend, or $r_i > x$ and the player does spend. The lower the cutpoint's score, the better it represents that person's decisions. In these two tables we report the data from one of the $\{6, 1\}$ treatments and one of the $\{6, 9\}$ treatments.[7]

The first thing to notice is that the minimum error cutpoint is not always unique. When forced to estimate a unique cutpoint, we select the one closest to 1, which is Nash play. In Table 9, subjects $\{4, 6, 10, 14, 15, 16\}$ are classified as Nash players, as are subjects $\{3, 4, 5, 6, 8, 10, 12, 13, 15\}$ in Table 10. A second thing to notice is that not every subject has the same estimated cutpoint. In Table 9, for example, subject #2 has an estimated cutpoint of 2.17 (corresponding to a token value of 13) while subject #16 has an estimated cutpoint of 1.0 (corresponding to a token value of 6). Another observation is that, for some subjects, the minimum number of errors is strictly greater than zero.

Pooling across all experiments, we find that 144/256, or 56 percent of the observations are Nash players. The entire distribution of cutpoints is illustrated in Figure 9. On the x-axis is the difference between the estimated cutpoint and the value of the public good in token value units. For example, subject #1 from Table 10 would be included in the "3" category in this figure, since his estimated cutpoint is 9 and the value of the public good is 6. An x-value of 0 in this figure corresponds to Nash play. This figure can also be broken down by experience, and doing so illustrates the effect of experience on inducing Nash (non-altruistic) play. This is shown in Figure 10.

Finally, we define *consistent players* as players that can be perfectly classified, so that they never make an error at their estimated cutpoint. Pooling across all experiments, we find 178/256, or 70 percent consistent players. The percentages of experienced and inexperienced consistent players are 75 and 64 respectively. Figure 11 displays the distribution of error rates, measured as the proportion of an individual's decisions that are inconsistent with his estimated cutpoint. Comparing to the earlier baselines, these error rates are again mostly in a range of five percent or below.

5.4 Comparison to Previous Results

There are a few simple comparisons between our data and the data from four person experiments conducted by Isaac and Walker. Recall that, in Issac and Walker's experiments, all subjects have identical marginal rates of substitution, equal to either 1.33 or

[6]When a particular rule was imprecise, *i.e.*, when the player was indifferent, it was assumed that no errors were made.

[7]These two tables are meant to be representative.

3.33 (which they refer to as High MPCR and Low MPCR). Their experiments also used a ten-period repetition design.

The most notable difference between their data and ours is in the frequency with which we observed *consistent Nash* play. This occurs when a subject, for an entire ten--period repetition, makes no decision that is inconsistent with dominant strategy Nash equilibrium. In terms of Figures 9 and 11, these subjects are in the 0-categories in *both* figures. We observe this 118 out of 256 observations, or 45 percent of the time. Isaac and Walker observe this 7 out of 76 observations, or 9 percent of the time. Thus we find five times as much consistent Nash play. Large differences also occur in the frequency of splitting, as pointed out earlier (Tables 2–5).

A second comparison is to look at the decisions made by our subjects when they had MRS = 1.33 and MRS = 3.33. The comparison is given in Table 11.

[Table 11 here]

Again, the same kind of pattern emerges. We find lower contribution rates. In fact, our contribution rate for MRS = 3.33 is roughly the same magnitude as the background noise measured in our baselines.

A third comparison is what we call repetition effects and what has been referred to elsewhere as *decay* — it is typical in these experiments to see less contribution in later periods than in early periods. In fact, in comparable experiments, contribution rates in early periods have ranged from two to four times as much as contribution rates in later periods. We measure an effect in our data (recall the Probit analysis), but we find the magnitude of the decay to be very small. It is true that there is more free riding in later periods, but this is attributable to a decrease in subject errors, or an increase in their consistency, not to a change in their decision rule. This fact is also reflected in the decline of splitting behavior documented earlier.

Andreoni (1988) conducted experiments similar to those of Isaac and Walker and observed magnitudes of contribution, free riding, and decay that by interpolation are roughly the same as those found in the data generated by Isaac and Walker. Those experiments used five person groups and MRS = 2. Instructions were somewhat different and some new treatments were explored. Andreoni's results are similar to those of Issac and Walker, and differences between our data and his are likewise similar to the differences between our data and Issac and Walker's.

Our findings also contrast sharply with the highly anomalous behavior in the experiments done by Saijo and Yamaguchi. They conducted homogeneous preference experiments with MRS = .7 and MRS = 1.42. Like Andreoni, they observe magnitudes of free riding, and decay for their experiments with an MRS = 1.42 that are roughly the same as those in Isaac and Walker's data. Saijo and Yamaguchi and Issac and Walker also observe similar split rates. The splitting rates observed in *both* of Saijo and Yamaguchi's treatments are 55 percent. They get as much splitting when subjects have a dominant strategy to contribute, as when subjects have a dominant strategy to free ride! Our findings are *dramatically* different.

Saijo and Yamaguchi observe aggregate contribution rates that are different from ours and also from Isaac and Walker's. For the 1.42 treatment, they observe 27 percent contribution, which is quite a bit less contribution than that seen in Isaac and Walker's data for MRS = 1.33. Our closest observations to MRS = 1.42 are at MRS = 1.5 and MRS = 1.4. We observed contribution rates of .27 and .36, respectively for those two values of MRS.

In their MRS = .7 treatment, Saijo and Yamaguchi see a contribution rate of 58 percent! Recall that our observed contribution rate was so close to 1 (.96) for this range of MRS, that we used this as one of our baselines for the rate of background noise! We have no satisfactory explanation for this enormous difference between their results and ours. However, we do note that those experiments were conducted somewhat differently in a number of ways, which may partially account for the differences in data.

Saijo and Yamaguchi employed seven member groups instead of four member groups, they conducted the experiments manually instead of through a computer network, and they used different instruction methods. In fact, they used two instruction sets as a treatment, and found significant differences due to that treatment. Also, they required subjects to make each decision within 20 seconds, and they used a different subject pool. Saijo and Yamaguchi suggest that the differences may be attributable to cultural differences between Japan and the U.S. We are skeptical of that explanation, but have no better one to offer.

6 Interpreting the Results

The main differences between our findings and previous findings can be summarized by the following observations:

1. We observe less splitting.
2. We do not observe significant decay.
3. We observe lower contribution rates.
4. We observe more Nash behavior.
5. We observe essentially no spiteful behavior.

The findings that replicate from past experiments with comparable group sizes are that experience leads to lower contribution rates, and contribution rates are declining in the marginal rate of substitution (marginal valuation of the private good).

Explanations for the differences that we observe are either methodological or environmental in nature. Possible methodological explanations abound: we utilize slightly different experimental procedures, or our instructions and computer screens are different, we employ a different subject pool, *etc.* On the environmental side, our experiments utilize a different economic environment, by which we mean the information structure and the profile of preferences in the group are different. In particular, as emphasized in the introduction, the information structure and profile of preferences correspond almost exactly to the standard environment used for auction experiments. In each period, preferences in the group are randomly and independently drawn from a known distribution of marginal rates of substitution, thereby inducing heterogeneity across individuals. This contrasts sharply with environments that have been explored in earlier investigations of the voluntary contributions mechanism.

To try to assess the relative importance of the methodological and environmental explanations, we have subsequently tried to replicate Issac and Walker's findings using our procedures and subject pool and their homogeneous environment. Specifically, we conducted an additional experimental session where every subject had a publicly announced marginal rate of substitution equal to 3.33, and every subject was endowed with multiple units of the private good.

Figure 12 compares the results of this session with the data from Issac and Walker. There is very little difference. The main features of the data replicate: there are very high contribution rates early on, and these rates decay significantly. In this extra session, we also observed similar splitting rates and amounts of Nash behavior. Based on this data, we dismiss the possibility that differences in our experimental procedures or subject pool are responsible for the differences in our results.

Thus we are left only with environmental explanations. This leads us to conclude that the findings from earlier experiments, experiments that utilized homogeneous environments, are not robust to public goods environments which exhibit variation in preferences, even if we limit attention only to linear public goods environments. This is a significant finding, even more so if one suspects, as we do, that heterogeneous preferences are a factor in most natural settings. There is an interesting question left open, namely "Why does heterogeneity lead to such different results?"

It is possible that, with homogeneous preferences, it is easier for a group to achieve a cooperative solution of the sort suggested by repeated game arguments. For example, if subjects adopt the type of strategies that reciprocate generous behavior by others, or believe that others adopt these strategies (see Kreps, Milgrom, Roberts, and Wilson [1985]), then some of the patterns of behavior that have been noticed in the homogeneous preference experiments, decay and pulsing, for example, can be rationalized.

In our design, since preferences are private information, the ability to signal one's generosity to other players is interfered with.[8] If one is observed to contribute, other

[8]Actually, in most of the homogeneous design experiments, homogeneity is not publicly announced. However, experiments by Isaac and Walker (1990) find that common knowledge of the homogeneity has

subjects cannot tell if you are being generous, or simply acting selfishly.

To identify the effects of the private information in our experiments, we conducted two revealed-information sessions (with $V = 6$ and $X = 9$) where all token value draws were revealed to everyone in the group. In the first of these sessions, token values were revealed after the decisions were made. In the second, token values were revealed before the decisions were made. In both cases, the signal interference problem is eliminated, which, if the above explanation is correct, should lead to greater contribution and less free riding.

The pooled results for the revealed information sessions are displayed in Figure 13, which compares the empirical response function with the data from all the other heterogeneous preference experiments (those with no revealed information).[9] There is very little difference. In fact, if anything, revealed information seems to lead to even more free riding behavior, which is contrary to the reputation hypothesis.

This leaves us without a complete explanation for why we observe such different results in our environment. At this point, we simply do not know. A number of other possible explanations can be imagined. Perhaps it was important (because of faster learning, less boredom, or something else) that subjects in our design are assigned a new MRS for each decision. This sort of explanation unfortunately seems to be currently beyond the reach of existing theoretical models of behavior in these kinds of games. On the other hand, the findings here are suggestive of possible new directions for theoretical work, as well as some directions for new experimental designs.

no effect on behavior. They conjecture that subjects infer from the wording in the instructions that other subjects have similar payoff tables.

[9]There is no significant difference between the two revealed information sessions, so pooling the data is reasonable.

Appendix

Sample Instructions from 4/9/92 (read aloud)

This is an experiment in decision making. You will be paid IN CASH at the end of the experiment. The amount of money you earn will depend upon the decisions you make and on the decisions other people make. It is important that you do not talk at all or otherwise attempt to communicate with the other subjects except according to the specific rules of the experiment. If you have a question, feel free to raise your hand. One of us will come over to where you are sitting and answer your question in private.

This session you are participating in is broken down into a sequence of four separate experiments. Each experiment will last 10 rounds. At the end of the last experiment, you will be paid the total amount you have accumulated during the course of all 4 experiments. Everyone will be paid in private and you are under no obligation to tell others how much you earned. Your earnings are given in FRANCS. At the end of the last experiment, you will be paid 11 cents for every 100 FRANCS you have accumulated during the course of all 4 experiments.

In each experiment you will be divided into 4 groups of 4 persons each. Those groups will stay the same for all 10 rounds of the experiment. After each of 10 round experiment, everyone will be regrouped into 4 entirely new groups. Therefore, whenever we change groups, the other people in your group will be completely different from the last group you were in. You will not be told the identity of the other members in your group. Since we will be running 4 experiments tonight, you will be assigned 4 different groupings, one for each 10 round experiment.

RULES FOR EXPERIMENT #1

Each round of the experiment you will have 9 tokens. You must choose how many of these tokens you wish to keep and how many tokens you wish to spend. The amount of money you earn in a round depends on how many tokens you keep, how many tokens you spend, and how many tokens are spent by others in your group. Each round, you will be told how many FRANCS each token is worth if you keep it. This amount, called your TOKEN VALUE, and will change from round to round and will vary from person to person randomly. To be more specific, in each round, this amount is equally likely to be anywhere from 1 to 20 FRANCS. There is absolutely no systematic or intentional pattern to your token values or the token values of anyone else. The determination of token values across rounds and across people is entirely random. Therefore, everyone in your group will generally have different token values. Furthermore, these token values will change from round to round in a random way. You will be informed PRIVATELY what your new token value is at the beginning of each round and you are not permitted to tell anyone what this amount is.

After being told your token value, you must wait at least 10 seconds before making your decision of how many tokens to spend and how many to keep. Your keyboard will be frozen for this period of time. When everyone has made a decision, you are told how many tokens were spent in your group and what your earnings were for that round.

This will continue for 10 rounds. Following each round you will begin with 9 new tokens and you will be randomly assigned a new token value between 1 and 20 FRANCS.

PAYOFFS

You will receive 3 FRANCS times the total number of tokens spent in your group. In addition, you will also receive your token value times the number of tokens you keep. Notice that this means every time anyone in your group spends a token, everyone in the group (including the spender) gets an additional 3 FRANCS, but the spender forgoes his or her token value for that token. WHAT HAPPENS IN YOUR GROUP HAS NO EFFECT ON THE PAYOFFS TO MEMBERS OF THE OTHER GROUPS AND VICE VERSA. Therefore, in each round, you have the following possible earnings, as shown in the table:

[WRITE EARNINGS TABLE ON BOARD]

Earnings Table for Experiment 1

YOUR SPENDING DECISION	OTHERS	YOUR EARNINGS (in FRANCS)
0	N tokens	(N*3) + (9*your token value)
1	N tokens	3 + (N*3) + (8*your token value)
2	N tokens	6 + (N*3) + (7*your token value)
3	N tokens	9 + (N*3) + (6*your token value)
4	N tokens	12 + (N*3) + (5*your token value)
5	N tokens	15 + (N*3) + (4*your token value)
6	N tokens	18 + (N*3) + (3*your token value)
7	N tokens	21 + (N*3) + (2*your token value)
8	N tokens	24 + (N*3) + your token value
9	N tokens	27 + (N*3)

Here is an example:

Suppose everyone else in your group spends 13 tokens in all and you spend 4 tokens and your token value was 12. You would earn 12 + 39 + 60 = 111 FRANCS. If you had spent 3 tokens you would have earned 9 + 39 + 72 = 120 FRANCS. If you had spent 5 tokens you would have earned 15 + 39 + 48 = 102 FRANCS.

ADDITIONAL PROCEDURES:

Are there any questions? [ANSWER QUESTIONS]

[Two practice rounds. Tell them not to press any keys unless you tell them to. In round 1 have each subject spend the number of tokens equal to the last digit of their ID#. In round 2 have each subject KEEP the number of tokens equal to the last digit of their ID#. Go over screen display and history. Tell subjects to refrain from pressing keys for no reason.]
[Keep screen display on]
[Hand out quiz.]
[Correct quiz answers and read them aloud.]
[Answer any additional questions.]

[Begin experiment 1.]

Specific instructions for Experiment 2:

Experiment 2 is the same as experiment 1 except you now have been regrouped with a completely different set of participants.

[Begin experiment 2.]

Specific instructions for Experiment 3:

Experiment 3 is the same as experiments 1 and 2 except now everyone in a group receives 15 FRANCS times the number of spenders in their group. Again, in addition, nonspenders also receive their token values. Again, everyone has been reassigned to a new group with a new set of participants. Here is your new payoff table:

[CHANGE BOARD. EXPLAIN.]

Earnings Table for Experiment 3

YOUR SPENDING DECISION	OTHERS	YOUR EARNINGS (in FRANCS)
0	N tokens	(N*15) + (9*your token value)
1	N tokens	15 + (N*15) + (8*your token value)
2	N tokens	30 + (N*15) + (7*your token value)
3	N tokens	45 + (N*15) + (6*your token value)
4	N tokens	60 + (N*15) + (5*your token value)
5	N tokens	75 + (N*15) + (4*your token value)
6	N tokens	90 + (N*15) + (3*your token value)
7	N token	105 + (N*15) + (2*your token value)
8	N tokens	120 + (N*15) + your token value
9	N tokens	135 + (N*15)

Example:
Suppose everyone else in your group spends 13 tokens in all and you spend 4 tokens and your token value was 12. You would earn 60 + 195 + 60 = 315 FRANCS. If you had spent 3 tokens you would have earned 45 + 195 + 72 = 312 FRANCS. If you had spent 5 tokens you would have earned 75 + 195 + 48 = 318 FRANCS.

[Begin experiment 3.]

Specific instructions for Experiment 4:

Experiment 4 is the same as experiment 3 except you have been regrouped again.

[Begin experiment 4.]

Endowment \ V	3	6	10	15
1 token	2	2	2	2
9 tokens	2	2	2	2

Table 1: Each cell has two 10-period sequences of a cohort with sixteen subjects divided into four groups. The first sequence is called "inexperienced"; the second is called "experienced." Groups were shuffled between sequences.

	early	late
inexp.	.22 (320)	.11 (320)
exp.	.12 (320)	.04 (320)

Table 2: Analysis of Splits. All data with endowment nine.

	early	late
inexp.	.36 (182)	.19 (176)
exp.	.21 (180)	.07 (170)

Table 3: Analysis of Splits. Endowment = 9, MRS > 1.

	early	late
inexp.	.029 (138)	.021 (144)
exp.	.021 (140)	.0067 (150)

Table 4: Analysis of Splits. Endowment = 9, MRS ≤ 1

	MRS = 1.33	MRS = 3.33
periods 1-5	.56 (120)	.60 (260)
periods 6-10	.56 (120)	.40 (260)

Table 5: Splitting behavior in the IW data.

	early	late
inexp.	.03 (262)	.04 (285)
exp.	.04 (263)	.04 (288)

Table 6: Spiteful behavior. Free-riding rates for subjects with MRS < 1 (Dominant Strategy to Contribute)

	early	late
inexp.	.08 (63)	.04 (65)
exp.	0 (65)	.002 (64)

Table 7: Sacrifical Behavior. Contribution Rates for Subjects with MRS > 4

Probit Models					
	1	2	3	4	5
ones	1.778	1.504	1.612	1.801	1.850
	(85.301)	(57.596)	(45.538)	(34.252)	(32.222)
MRS	-1.156	-0.916	-0.973	-1.013	-1.015
	(-86.358)	(-58.866)	(-44.078)	(-42.878)	(-42.896)
exper.s		-0.861	-0.858	-0.867	-0.868
		(-25.252)	(-25.084)	(-25.235)	(-25.233)
exper		0.983	0.980	0.992	0.994
		(20.013)	(19.919)	(20.075)	(20.089)
endow.s			0.104	0.108	0.107
			(3.742)	(3.888)	(3.856)
endow			-0.199	-0.207	-0.205
			(-4.618)	(-4.761)	(-4.730)
V				-0.015	-0.015
				(-4.923)	(-4.993)
period					-0.008
					(-2.146)
lg lkhd	-8912.7	-8522.7	-8511.9	-8499.7	-8497.4
% pred.	83.064	83.160	83.238	83.429	83.607

Table 8: In each Probit Model, the dependent variable is the investment decision. Equal weight has been given to both the one token treatment and to the nine token treatment. Under each coefficient is the asymptotic t-statistic. The log likelihood and the percentage correctly predicted are also given for each model.

Token Value (Cutpoint)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	1	5	5	5	3	3	3	3	3	2	2	0	0	0	0	1	1	2	4	4	5
	2	3	2	2	2	1	1	1	1	1	1	1	1	0	0	0	0	0	2	4	5
S	3	5	4	4	3	3	3	3	3	2	2	1	1	1	2	3	3	3	3	3	4
u	4	2	1	1	1	0	0	0	0	0	0	0	1	3	3	3	4	5	5	5	5
b	5	4	3	2	2	2	2	1	1	0	0	0	0	1	4	4	4	4	5	5	5
j	6	2	2	2	2	1	0	0	0	1	2	2	4	4	4	5	5	7	7	7	7
e	7	3	3	2	2	1	1	1	0	0	2	2	2	2	2	3	5	5	5	5	7
c	8	4	4	3	3	3	3	1	0	0	1	2	2	2	2	2	3	3	4	6	6
t	9	5	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	4	4	4	4
	10	3	3	2	2	1	0	0	0	1	1	1	1	3	5	5	5	7	7	7	7
#	11	3	3	3	2	1	1	1	0	0	0	0	3	4	4	4	5	5	5	7	7
	12	5	5	4	4	4	4	3	2	1	0	1	1	1	2	3	4	4	4	4	4
	13	6	6	5	4	4	3	3	3	0	0	0	0	1	1	1	2	2	2	2	2
	14	2	1	1	1	0	0	0	1	3	4	4	5	6	6	6	7	7	7	7	8
	15	2	2	1	1	0	0	0	1	1	1	1	1	2	4	5	6	6	6	6	7
	16	3	2	1	1	1	1	1	2	2	2	2	3	4	4	4	5	6	6	7	7

Table 9: The raw number of classification errors for the first repetition of treatment $\{6,1\}$

Token Value (Cutpoint)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	1	23	23	23	5	5	5	5	5	3	10	8	24	24	23	31	31	40	58	58	67
	2	22	15	17	17	12	16	16	16	16	16	16	16	11	15	15	15	10	23	41	50
S	3	54	45	45	27	18	18	18	18	18	27	27	36	36	36	36	36	36	36	36	36
u	4	18	9	9	9	0	0	0	0	0	0	0	9	27	27	27	36	45	45	45	45
b	5	18	9	0	0	0	0	0	9	9	18	18	18	27	54	54	54	54	63	63	63
j	6	18	18	18	18	9	0	0	0	9	18	18	36	36	36	45	45	63	63	63	63
e	7	31	31	22	22	13	13	13	10	10	22	22	22	22	19	24	41	41	41	41	59
c	8	11	11	2	2	2	2	2	20	29	37	45	45	45	45	45	54	54	62	79	79
t	9	30	12	12	12	12	12	12	12	12	12	12	12	9	24	24	19	41	39	46	53
	10	27	27	18	18	9	0	0	0	9	9	9	9	27	45	45	45	63	63	63	63
#	11	18	18	18	18	18	18	18	14	19	19	14	36	45	45	45	54	54	54	72	72
	12	9	9	0	0	0	0	0	9	18	27	45	45	45	54	63	72	72	72	72	72
	13	19	19	10	1	1	1	10	10	9	35	35	35	44	44	44	53	53	53	53	53
	14	44	35	35	35	26	26	17	9	19	28	20	21	30	30	29	37	37	37	37	46
	15	22	22	13	13	4	4	4	13	13	13	13	13	20	34	41	50	50	50	50	59
	16	30	12	12	21	21	21	19	26	26	26	21	23	30	30	28	34	42	42	51	51

Table 10: The raw number of classification errors for the first repetition of treatment $\{6,9\}$

	IW data	Our data
MRS = 1.33	.50 (240)	.37 (90)
MRS = 3.33	.20 (520)	.05 (56)

Table 11: Contribution rates. Comparison to IW data, when MRS = 1.33 and MRS = 3.33

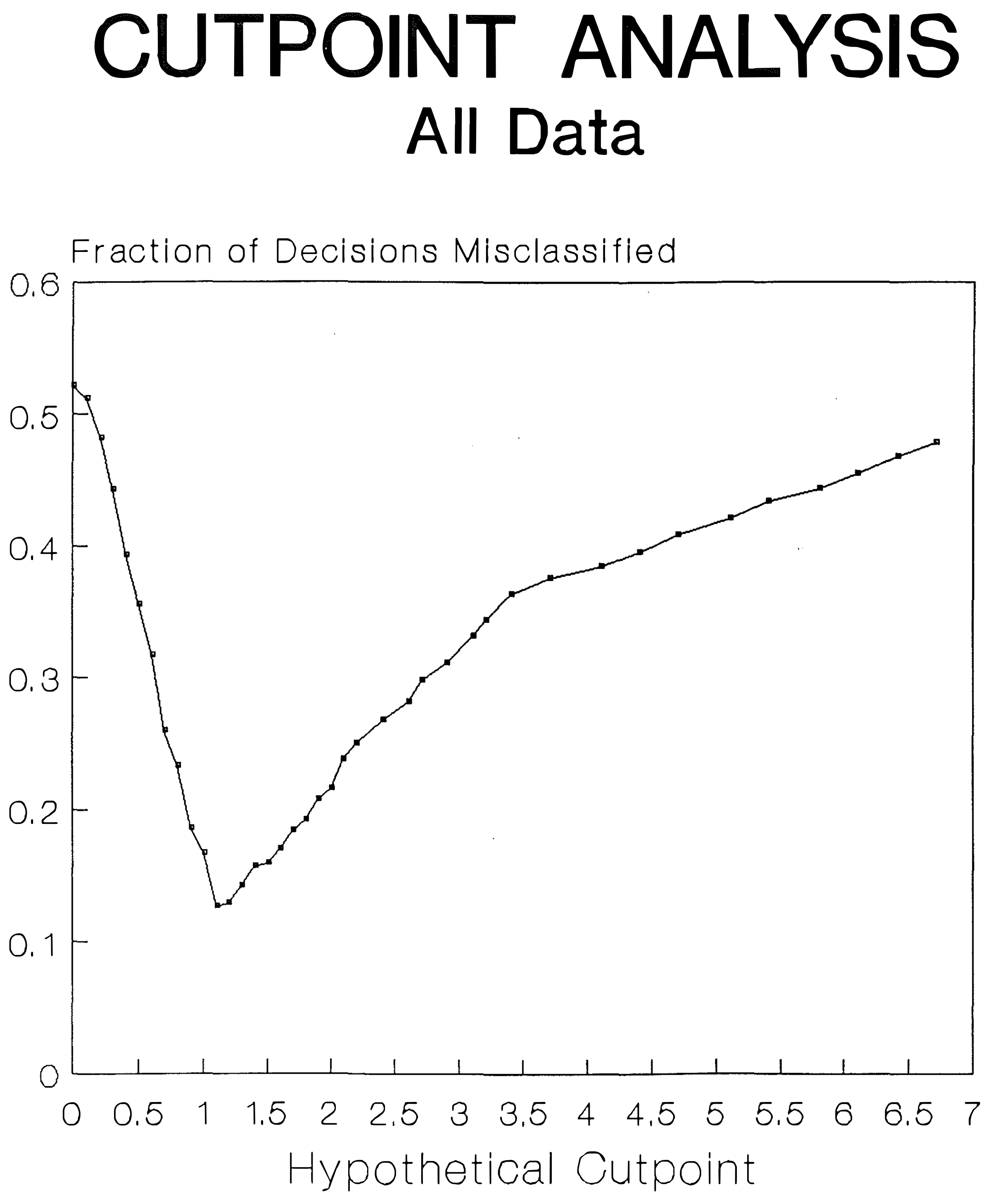


Figure 1: Cutpoint analysis: aggregate level

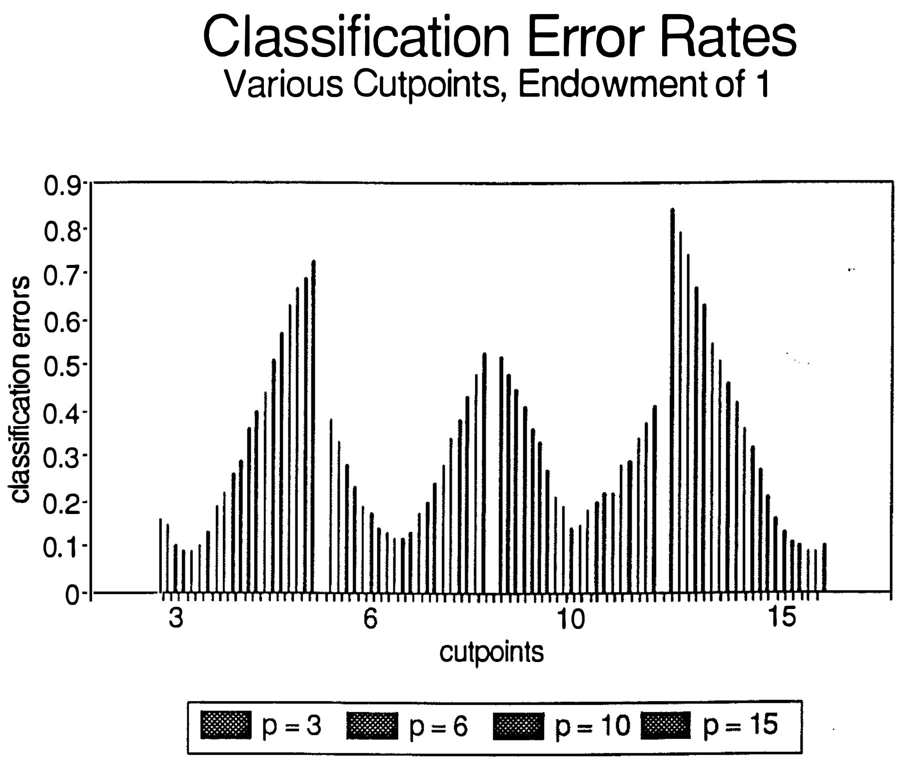


Figure 2: Classification errors aggregated over all subjects shown for all treatments with an endowment condition of one.

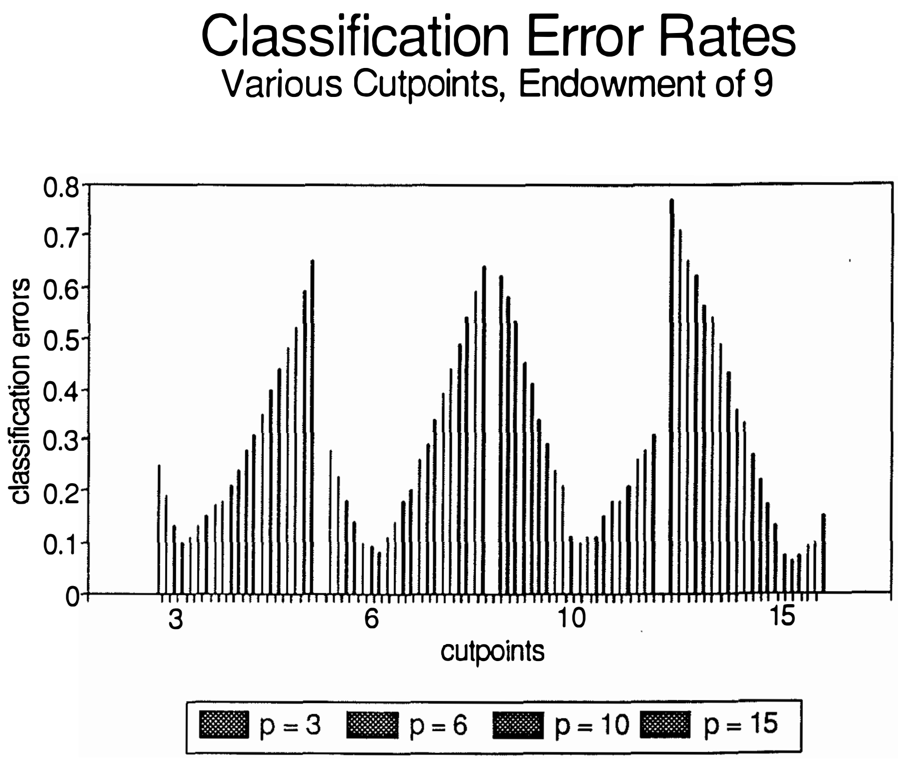


Figure 3: Classification errors aggregated over all subjects shown for all treatments with an endowment condition of nine.

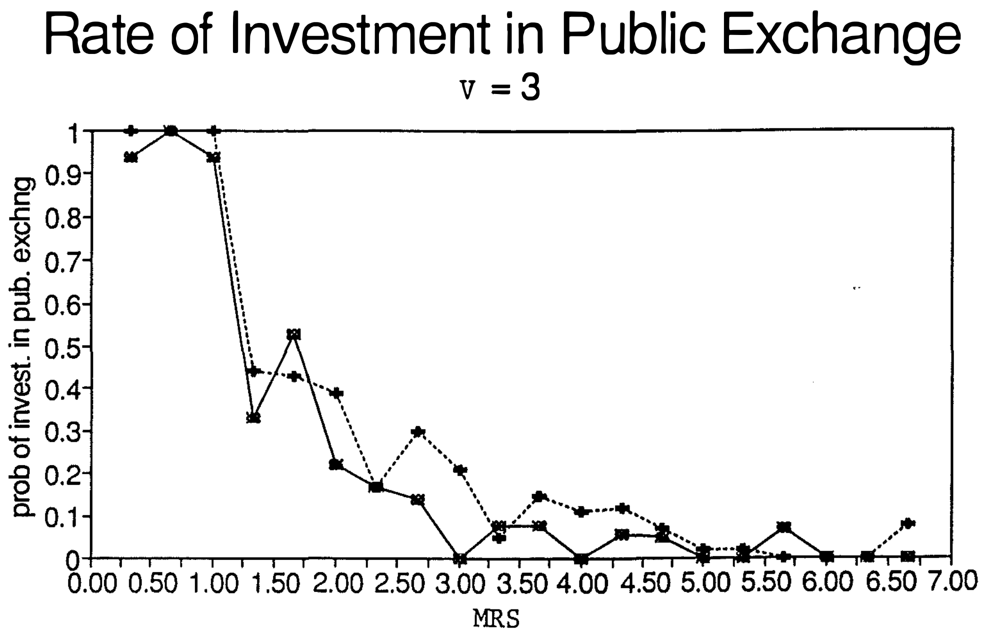


Figure 4: The aggregate percentage of tokens invested in the public exchange *vs.* the marginal rate of substitution, plotted for both the endowment of one and the endowment of nine conditions. $V = 3$

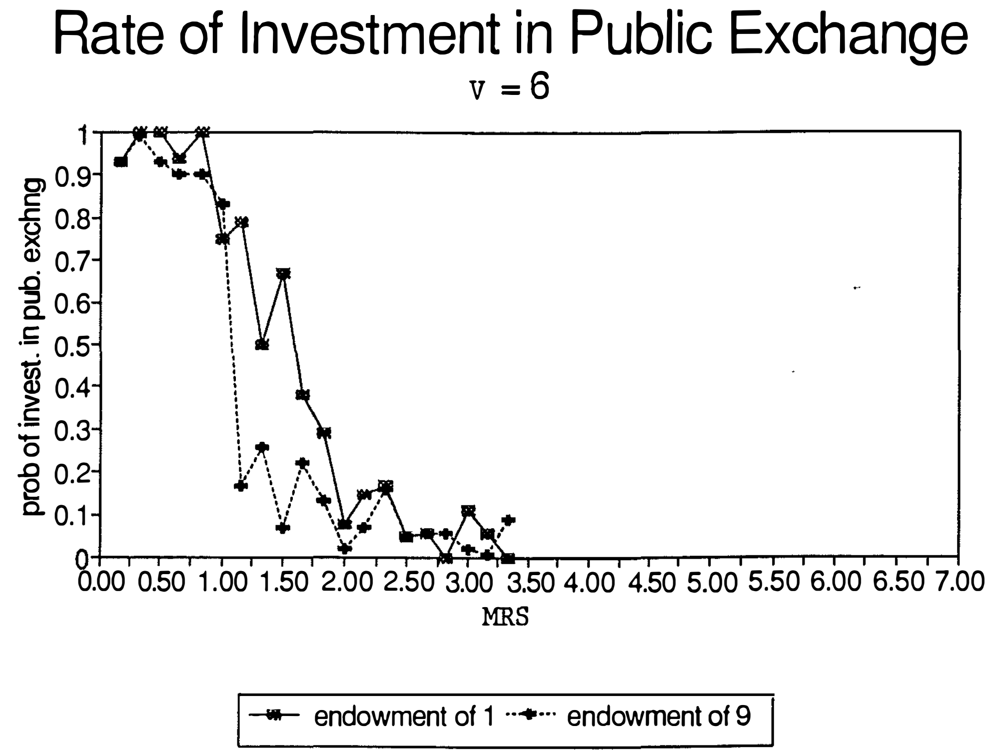


Figure 5: The aggregate percentage of tokens invested in the public exchange *vs.* the marginal rate of substitution, plotted for both the endowment of one and the endowment of nine conditions. $V = 6$

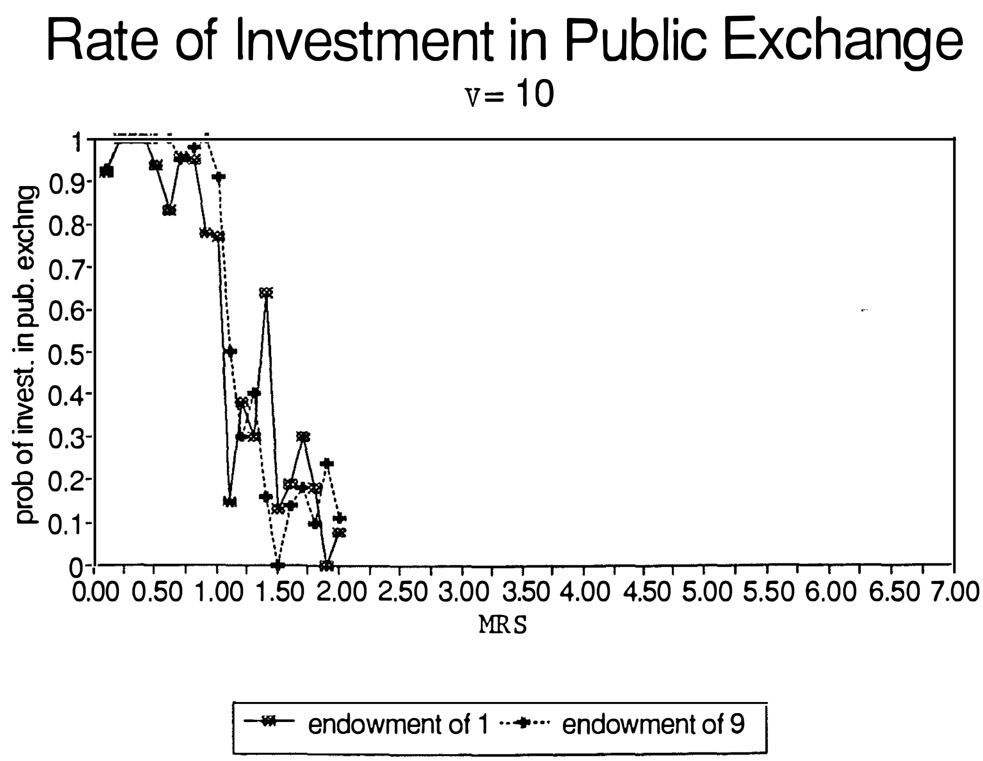


Figure 6: The aggregate percentage of tokens invested in the public exchange *vs.* the marginal rate of substitution, plotted for both the endowment of one and the endowment of nine conditions. $V = 10$

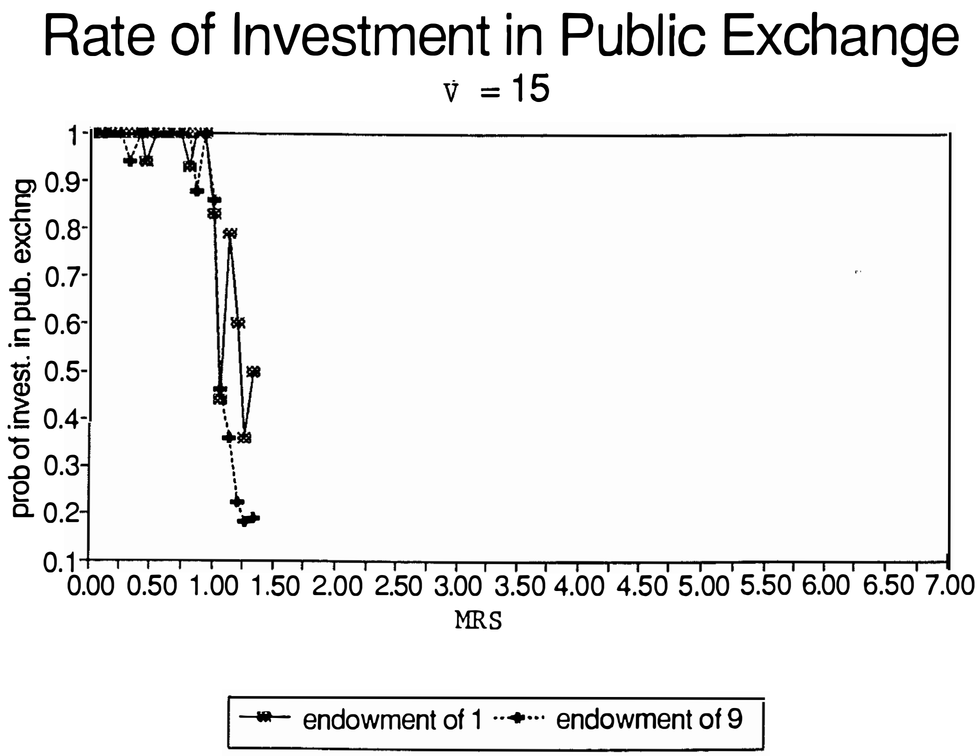


Figure 7: The aggregate percentage of tokens invested in the public exchange *vs.* the marginal rate of substitution, plotted for both the endowment of one and the endowment of nine conditions. $V = 10$

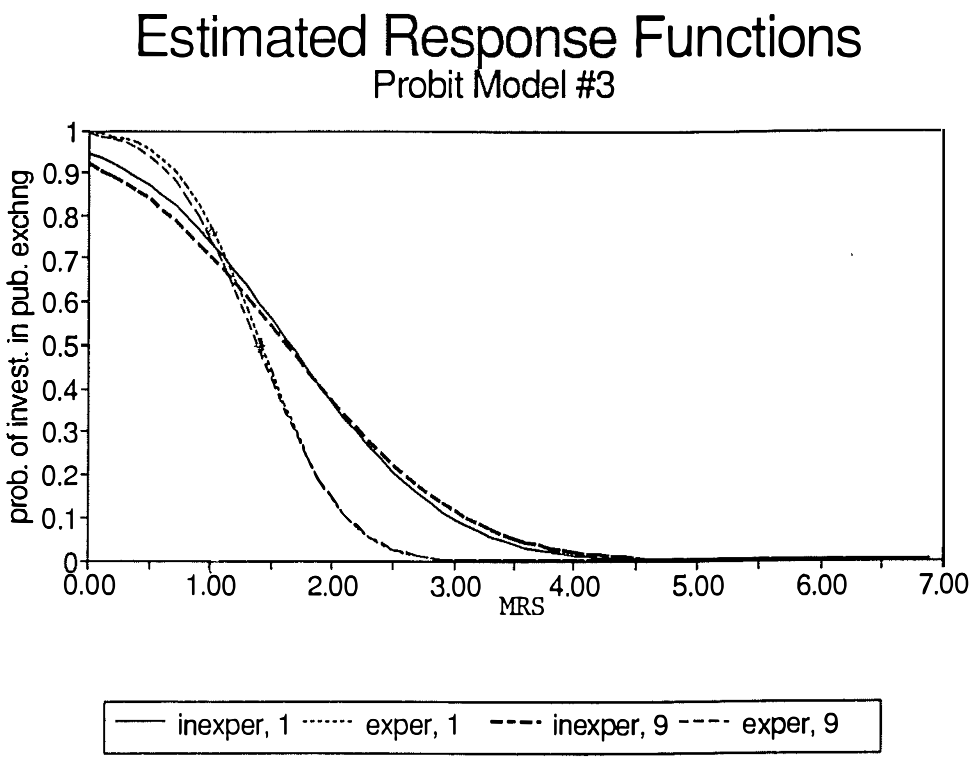


Figure 8: The different response functions generated by Probit Model No. 3.

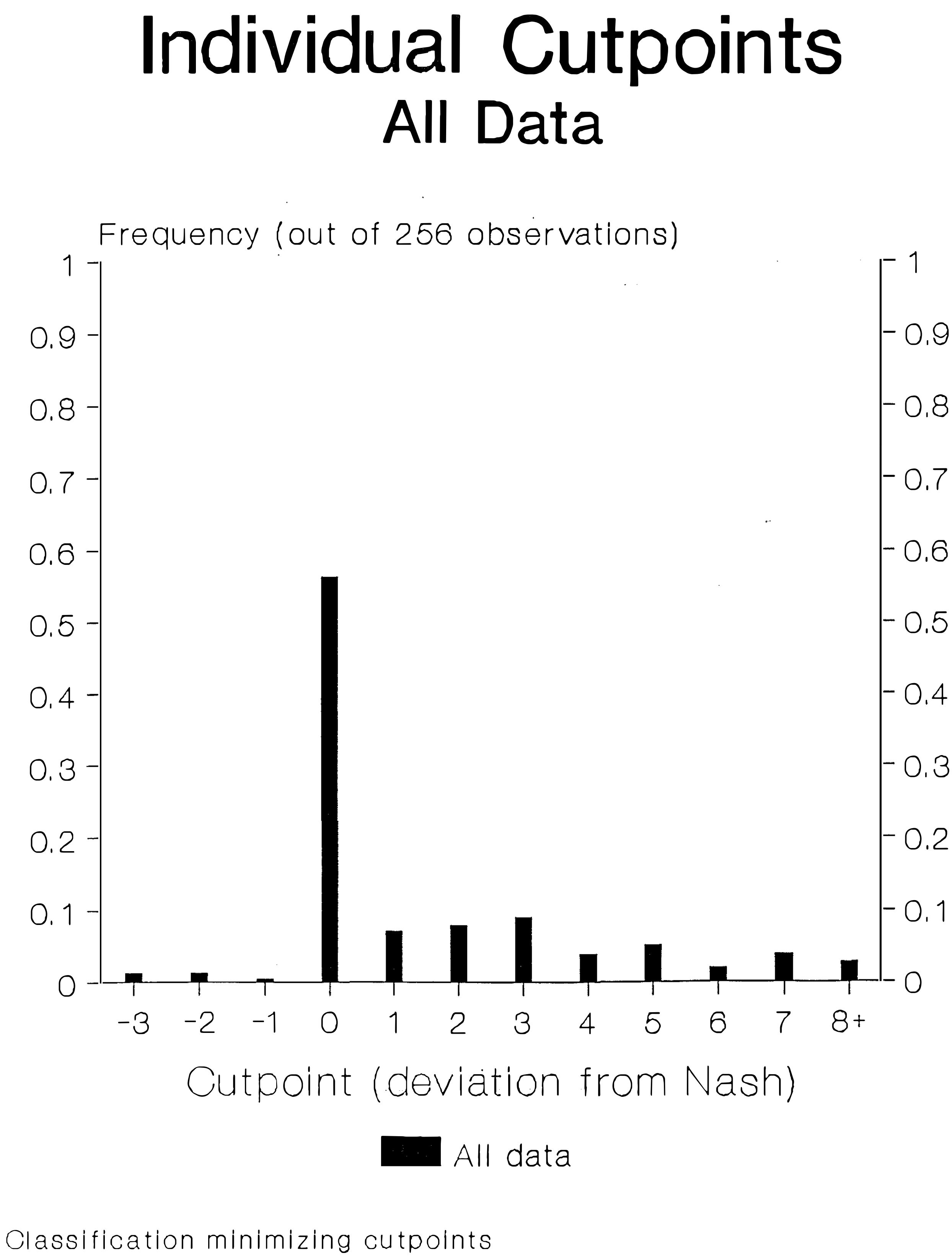
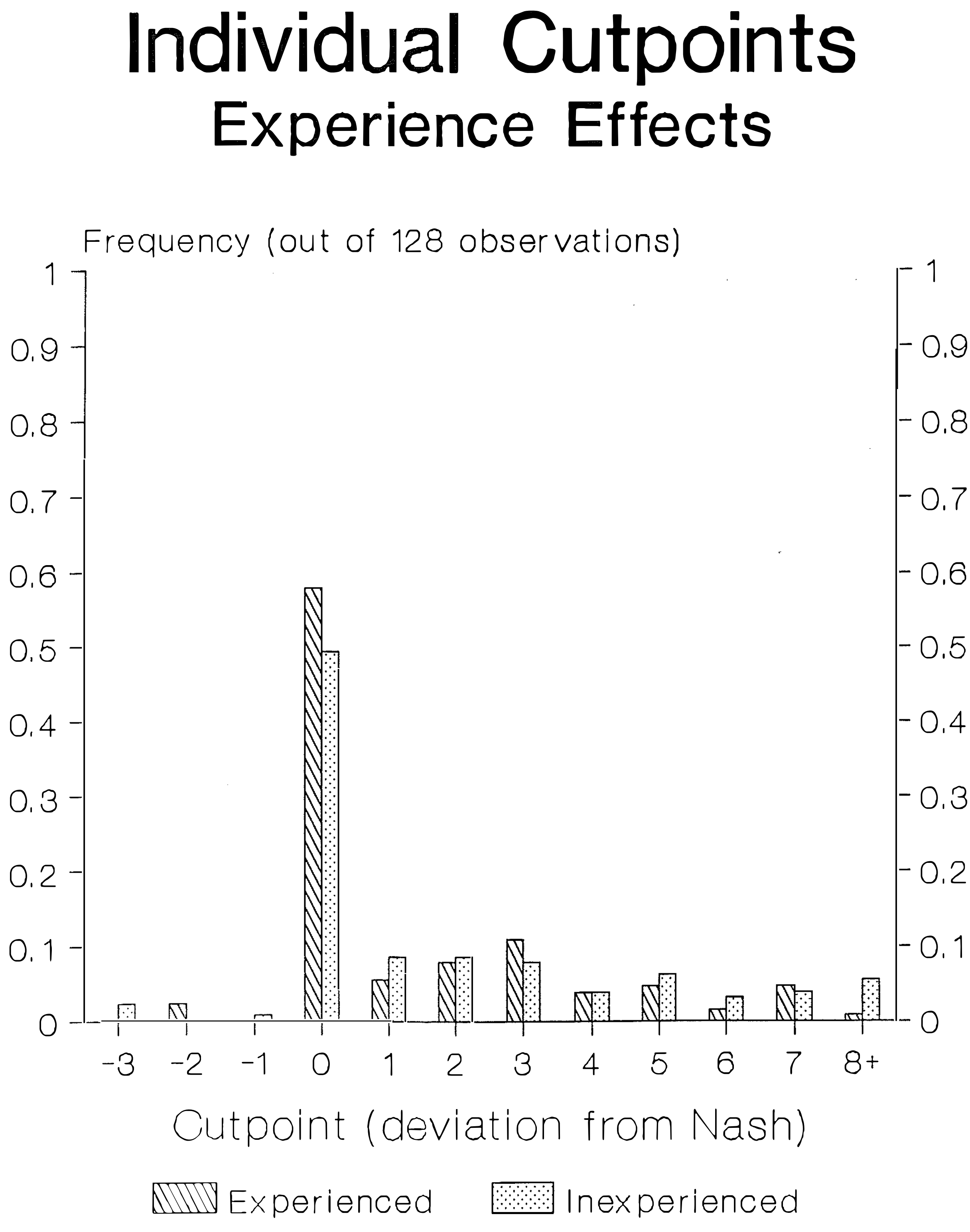


Figure 9: Estimated cutpoints measured as deviation from Nash play (in token value units). All data



Classification minimizing cutpoints

Figure 10: Estimated cutpoints measured as deviation from Nash play (in token value units). Experience effects

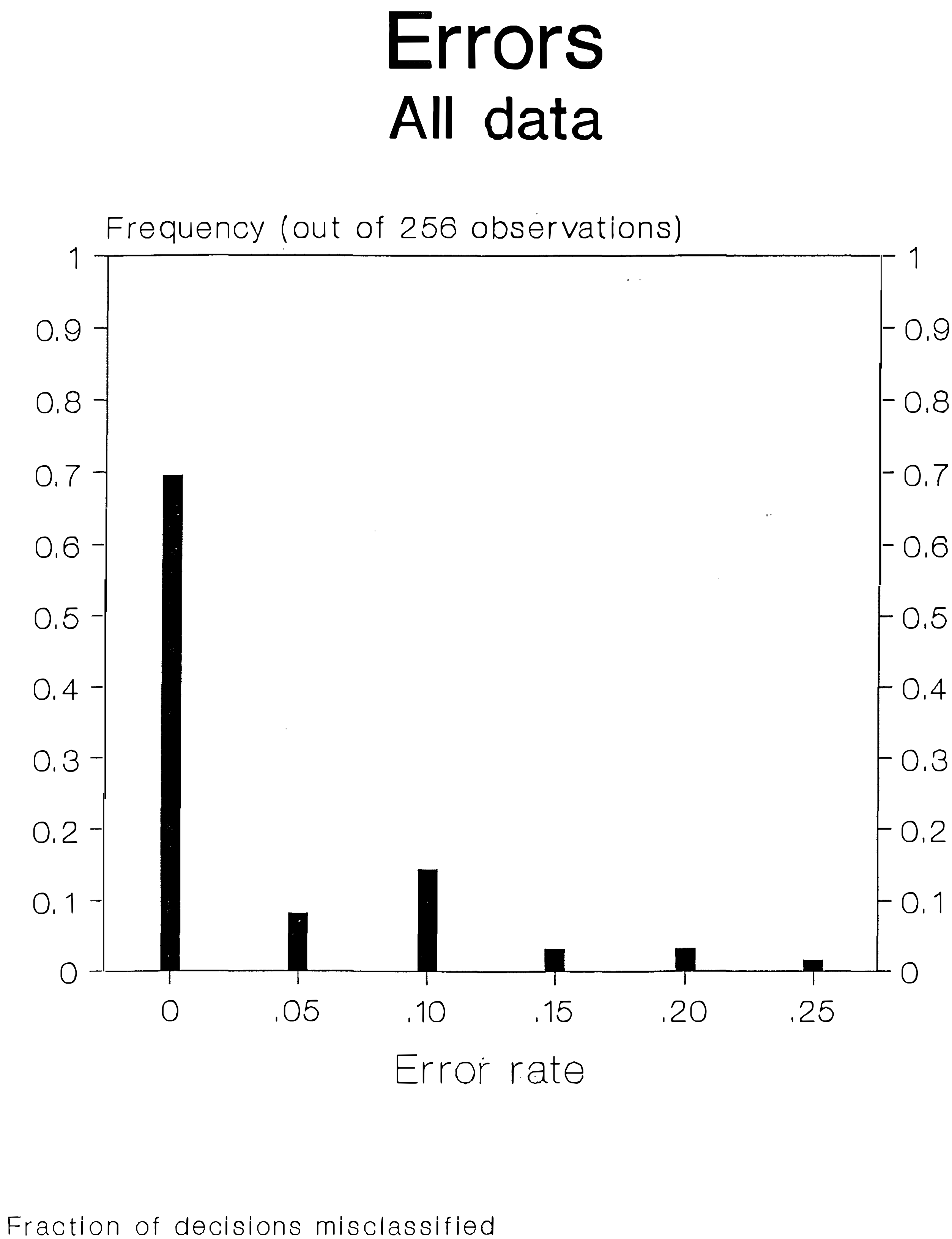


Figure 11: Classification errors.

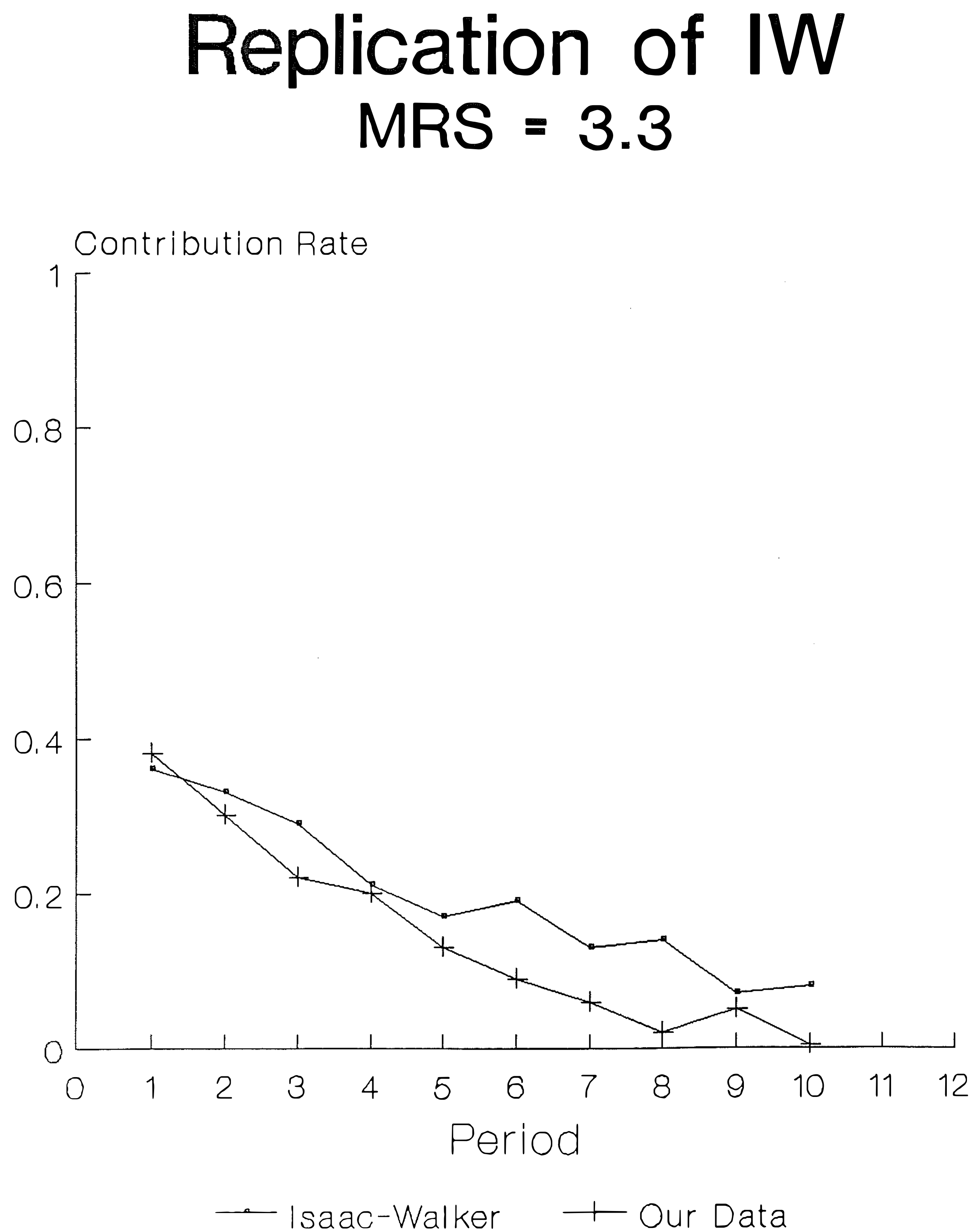


Figure 12. Replication of homogeneous preference experiments with $V = 6, r = 20, X = 9$ (MRS= 3.3)

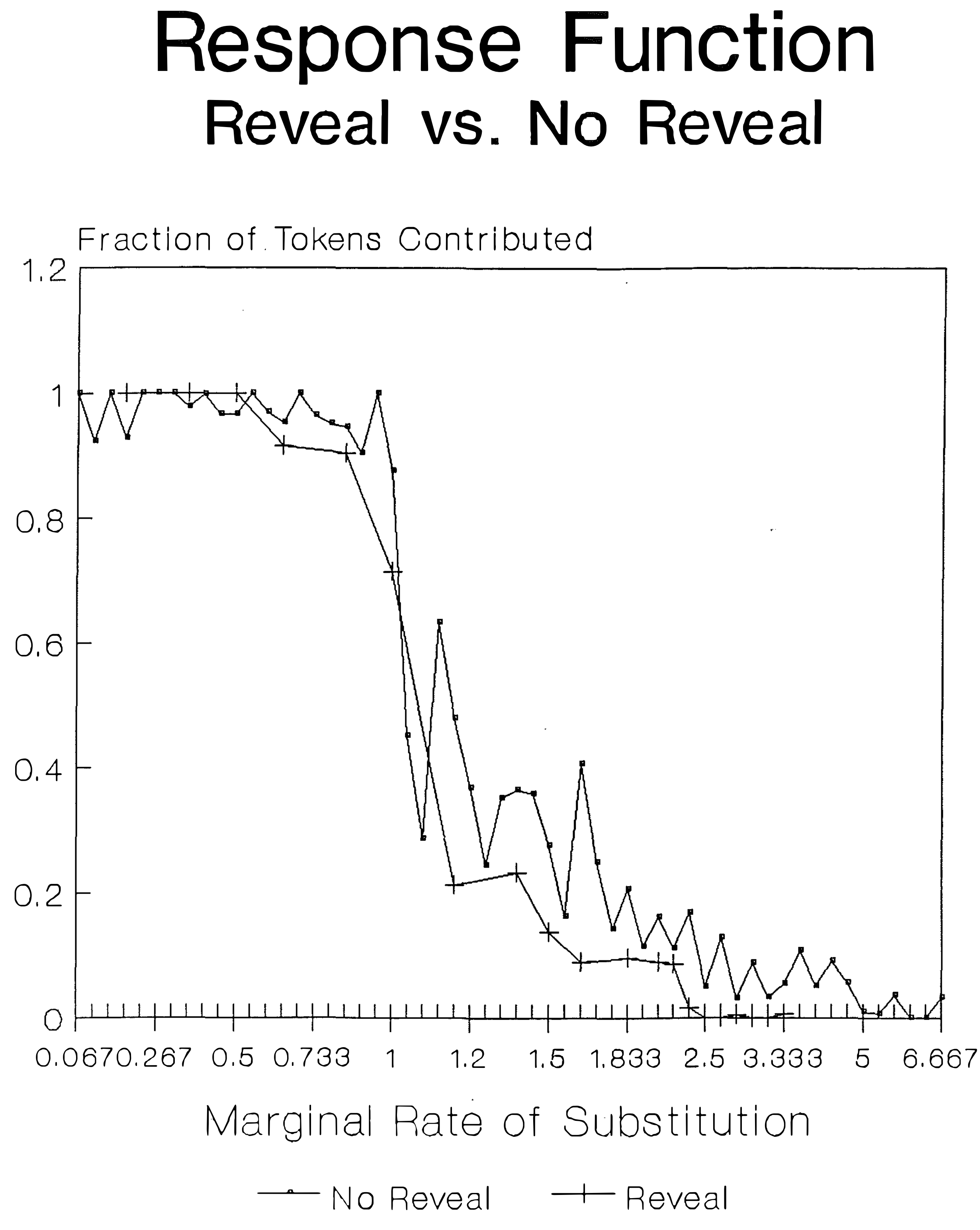


Figure 13. Empirical response function with (reveal) and without (no reveal) publicly reported token values.

References

[1] Andreoni, James. 1988. "Why Free Ride? Strategies and Learning in Public Goods Experiments." *Journal of Public Economics* Vol. 37: pp. 291 - 304.

[2] Andreoni, James. 1992. "Cooperation in Public Goods Experiments: Kindness or Confusion?" typescript.

[3] Brookshire, D. S. , D. L. Coursey, and D. B. Redington, 1989. "Special Interests and the Voluntary Provision of Public Goods," University of Wyoming Working Paper.

[4] Dawes, R. M. 1980. "Social Dilemmas." *Annual Review of Psychology* Vol. 31: pp. 169 - 193.

[5] Devore, J. L. 1982. *Probability and Statistics for Engineering and the Sciences*. Monterey, California: Brooks/Cole Publishing Company.

[6] Fisher, Joseph, R. Mark Isaac, Jeffrey W. Schatzberg, and James M. Walker. October 1991. "Heterogeneous Demand for Public Goods: Effects on the Voluntary Contribution Mechanism." unpublished manuscript.

[7] Isaac, R. Mark, Kenneth F. McCue and Charles R. Plott. 1985. "Public Goods Provision in an Experimental Environment," *Journal of Public Economics* Vol 26: pp. 51 - 74.

[8] Isaac, R. Mark, and James M. Walker. February 1988. "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism." *The Quarterly Journal of Economics*: pp. 179 - 198.

[9] Isaac, R. Mark, and James M. Walker, 1989. "Complete Information and the Provision of Public Goods," Discussion Paper 89-18, University of Arizona.

[10] Isaac, R. Mark, James M. Walker, and Susan H. Thomas. 1984. "Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations." *Public Choice* Vol. 43: pp. 113 - 149.

[11] Isaac, R. Mark, James M. Walker, and Arlington W. Williams. September 1991. "Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups." unpublished manuscript.

[12] Kim, Oliver, and Mark Walker. 1984. "The Free Rider Problem: Experimental Evidence." *Public Choice* Vol. 43: pp. 3 - 24.

[13] Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson, 1982. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 27:245-52.

[14] Ledyard, John, O. 1992. "Public Goods: A Survey of Experimental Research." unpublished manuscript, California Institute of Technology.

[15] Maddala, G. S. 1983. *Limited-dependent and Qualitative Variables in Economics*, Cambridge: Cambridge University Press.

[16] Marwell, Gerald and Ruth E. Ames. 1979. "Experiments on the Provision of Public Goods. I. Resources, Interest, Group Size, and the Free-Rider Problem," *American Journal of Psychology* Vol. 84, No. 6: pp. 1335 - 1360.

[17] Marwell, Gerald and Ruth E. Ames. 1980. "Experiments on the Provision of Public Goods II. Provision Points, Stakes, Experience and the Free Rider Problem." *American Journal of Psychology* Vol. 85: pp. 926 - 937.

[18] Marwell, Gerald and Ruth E. Ames. 1981. "Economists Free Ride, Does anybody Else? Experiments on the Provision of Public Goods, IV." *Journal of Public Economics* Vol. 15: pp. 295 - 310.

[19] McFadden, Daniel. 1982. "Econometric Models of Probabilistic Choice," in C. Manski and D. McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge, Mass.: MIT Press.

[20] Ostrom, E., J. Walker, and R. Gardner. 1991. "Covenants with and without a Sword: Self–Enforcement is Possible." Workshop in Political Theory and Policy Analysis Working Paper, Indiana University.

[21] Palfrey, Thomas R. and Howard Rosenthal. 1988. "Private Incentives in Social Dilemmas: The Effects of Incomplete Information and Altruism." *Journal of Public Economics* Vol. 35: pp. 309 - 332.

[22] Palfrey, Thomas R. and Howard Rosenthal. 1991. "Testing Game-Theoretic Models of Free Riding: New Evidence on Probability Bias and Learning." in *Laboratory Research in Political Economy* (T. Palfrey, ed.), Ann Arbor; University of Michigan Press.

[23] Saijo, T. and Y. Yamaguchi. 1992. "The "Spite" Dilemma in Voluntary Contributions Mechanism Experiments," unpublished manuscript.