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

Thomas R. Palfrey; Jeffrey E. Prisbrey

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

By Thomas R. Palfrey and Jeffrey E. Prisbrey *

We report the results of voluntary contributions experiments where subjects are randomly assigned different rates of return from their private consumption. These random assignments are changed round to round, enabling the measurement of individual player contribution rates as a function of that player's investment cost. We directly test these response functions for the presence of warm-glow and/or altruism effects. We find significant evidence for heterogeneous warm-glow effects that are, on average, low in magnitude. We statistically reject the presence of an altruism effect. (JEL C92, C92, H41)

There is a growing body of experimental data from voluntary contribution, public goods environments with a single public good and a single private good. Among the many features of the data that are difficult to explain is the apparent frequent use of strictly dominated strategies. Subjects not only give away money when free-riding is a dominant strategy (R. Mark Isaac et al. [1984, 1994], Isaac and James M. Walker [1988, 1989], and elsewhere), but they also often fail to contribute when it is in their own best interests to do so (Tatsuyoshi Saijo and Hideki Nakamura, 1995). Furthermore, individual behavior over time exhibits erratic patterns, with many subjects alternating back and forth between generosity and selfishness. John O. Ledyard's (1995) excellent survey documents these and several other anomalies.

The anomalies might be cause for a serious reexamination of the theory, as they signal trouble for current economic models of selfish behavior. However, the range of environments for which these experimental results have been reported is very narrow, and more importantly the designs employed make it difficult, if not impossible, to estimate the actual strategies underlying subject behavior. Our design, by changing both the information structure and the distribution of preferences, allows the estimation of strategies at both the group and the individual level. As a result, we are able to clearly identify the different sources of some of these anomalies. The different environment also provides a chance to see if previous anomalous findings are robust.

We employed the following basic design, which is a variation on the Voluntary Contributions Mechanism (Isaac et al., 1984). Each subject was given an endowment which could voluntarily be contributed toward a public good, or kept to be consumed as a private good. The consumption value of the public good depended linearly upon the total contributions of the group. All the subjects in a group had the same commonly known, marginal value for the public good. But, individual subjects were randomly assigned different marginal values for the private good from a commonly known distribution. In such a setup, subjects whose value for the private good is less than their value for the public good have a dominant strategy to contribute all of their endowment; subjects whose value for the private good is greater than their value for the public good have a dominant strategy to keep all of their endowment or to free ride. Subjects repeated the game several

times, each time being randomly reassigned a new value for the private good.

Specifically, our laboratory environment consists of N individuals, each endowed with w_i discrete units of a private good. The marginal rate of transformation between the public good and the private good is one-for-one, and individual monetary payoff functions are of the form: $U(x_i, x_{-i}) = V\Sigma_j x_j + r_i(w_i - x_i)$, where x_i is the individual's contribution. We refer to V as the marginal value of the public good, and it is the same for all individuals. The marginal value of the private good is r_i, and it is private information.

Essentially all of what we think we know about behavior in this game is based on experiments in which the marginal valuations of the private good are identical in all periods for all participants in the experiments. With one exception,[1] the private good valuations exceed the public good valuation, so all subjects have a dominant strategy to free ride. The central findings from these experiments are summarized below.

• Most players in this game violate their one-shot dominant strategy, with many contributing upwards of half their endowment. They do so even when the marginal valuation of the private good is three or more times that of the public good.

• As the marginal valuation of the private good gets closer to the marginal valuation of the public good, more violations of the dominant strategy are observed.

• Subjects can be roughly categorized according to their tendency to violate the dominant strategy.

• Violations of dominant strategies diminish both with repetition and with experience (playing a second sequence of games with a new group).

• Violations of dominant strategies to contribute, i.e., when $r_i < V$, appear to be as prevalent as violations of dominant strategies to free ride (Saijo and Nakamura, 1995).

Several possible explanations have been offered for why there is so much more cooperation than the standard theory predicts. The explanations that have thus far received the most attention are:

(a) altruistic preferences;
(b) warm-glow preferences;
(c) repeated game effects, including reputation building; and
(d) subject confusion.

The first two explanations are similar because they both suggest that subjects have a nonmonetary component in their utility function that is difficult for the experimenter to control, and that works in the opposite direction of the monetary incentive to free ride. By altruistic preferences we mean that a subject's utility is increasing, not only in his or her own payoff, but also in the total group payoff. Warm-glow preferences mean that the *act* of contributing, independent of how much it increases group payoffs, increases a subject's utility by a fixed amount.

At first blush, these two effects would appear to be the same, but in fact they are not. Unlike the warm-glow explanation, the altruism explanation predicts that increases in group size and/or in the value of the public good should have very large effects on contribution rates. The warm-glow explanation does not depend upon group size or the marginal value of the public good.

The latter two explanations, (c) and (d), are suggested by the tendency for contributions to decline with repetition and with experience. The declines may be consistent with learning or endgame effects.

It is possible that the typical act of contribution is motivated, perhaps to differing degrees, by each of these explanations. One purpose of these experiments is to accurately measure subject behavior in order to cleanly separate between these explanations and ascertain their relative importance. To do so requires major design innovations relative to the standard public goods experiment. In the typical past experiments, all subjects within a group had the same r_i; here different subjects have different r_i's.[2] In the past, all subjects

---

[1] Saijo and Nakamura (1995).

[2] There are a few exceptions, notably Isaac et al. (1985) and Joseph R. Fisher et al. (1995), both of whom consider environments with two incentive types. The latter provides subjects with identical information about other subjects' preferences as in parallel homogeneous preference experiments. The former has several other different features, including nonlinearities, and does not conduct any base-

usually had a dominant strategy to free ride, while here the subjects sometimes have a dominant strategy to contribute. In the past, subjects repeated the decision with the same incentives each period; here the subject's incentives change each period.

In earlier experiments, a subject who contributed because of confusion or decision error could not be differentiated from a subject who contributed because of altruism or warm glow. Because r_i was always bigger than V, subjects never had an incentive to contribute, and therefore every contribution could be called a decision error. Behavior motivated by altruism or a warm glow, although potentially observed, could not be separately identified. Furthermore, it was impossible even to observe noncontribution when $r_i < V$.

Thus there is an inherent limitation in past designs. In our design this problem is eliminated and contribution arising from confusion or decision error can be differentiated from contribution due to nonmonetary components of the utility function.

Thus, a key benefit of our design is that the resulting data allows the accurate and unbiased measurement of strategies — measurement that controls for the possibility of subject error. And, directly from the estimated strategies come estimates of the amount of altruism and warm glow in the individual utility functions. We can also check for the robustness of existing results to environments that include important features, such as diverse preferences and incomplete information, that are endemic to natural settings.

I. Experimental Design and Procedures

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

1. Each subject participates in four sequences of ten periods (one decision per period), each ten-period sequence with a different group of three other subjects.[3] The first two such sequences have the same value of V. The last two sequences also have the same value of V, but different from the value in the first two sequences. This allows us to identify experience effects. The first sequence with each value of V is coded as inexperienced, and the second sequence as experienced.[4] All four sequences occur in a single session that lasts approximately 90 minutes. Each session includes 16 subjects.

2. In all our environments, subjects receive r_i's that are randomly assigned according to a uniform distribution between 1 and 20 in unit increments. We sometimes refer to these as token values. Each time a subject is to make a new decision, he or she 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 of the experiment. The value of V is also publicly announced.

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

3. We vary the value of the public good, V, between experiments. We have an equal number of observations for each of the four different values of $V \in \{3, 6, 10, 15\}$ (see Table 1). One value, $V = 3$, has the feature that group efficiency is not maximized when all subjects contribute in every decision period. In that condition, on average, 40 percent of the time subjects are assigned a token value that is worth more than four times the individual

---

line experiments with homogeneous preferences. Gerald Marwell and Ruth E. Ames (1980) and D. S. Brookshire et al. (1989) have also conducted experiments with heterogeneous preferences, but these are not comparable for other reasons. None of these experiments varied individual incentives across decisions, nor did they provide explicit information about the distribution of incentives in the population. Palfrey and Howard Rosenthal (1991) use an environment similar to the one explored here, but the public good technology is step-level, not linear.

[3] Fixing the groups for a sequence of ten periods was done to maintain comparability with past experiments. We also conducted a replication of one of the Isaac et al. (1984) treatments, using our instructions, computer protocol, and subject pool. We obtained results, reported in Palfrey and Prisbrey (1993), that were similar to Isaac et al. (1984).

[4] Alternative ways of coding experience produce similar results.

TABLE 1—SESSION NUMBER AND SEQUENCE NUMBERS
FOR EACH OF THE EIGHT TREATMENTS

			V		
Endowment		3	6	10	15
1 token	Session #	1	3	1	3
	Sequence #'s	1, 2	1, 2	3, 4	3, 4
9 tokens	Session #	2	4	4	2
	Sequence #'s	1, 2	1, 2	3, 4	3, 4

Notes: The experiment consisted of four sessions, each with four ten-period sequences. This table indicates session number and sequence numbers for each of the eight treatments.

marginal value of the public good. In these cases, contribution *reduces* group efficiency.

4. 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, and can contribute any number between zero and nine in each period (see Table 1).

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

6. Each subject was paid cash for each point he or she earned in the session. On average, each individual subject earned approximately $15 in a session.

II. Data Analysis

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, the analog to bidding functions in auctions. The response functions address the question: How do contribution decisions depend on the private token values and the public good value, and how do these functions change with our treatment variables, such as experience? We estimate response functions at both the aggregate and individual levels, using probit models.

One can interpret our analysis in the context of a random utility model, of the sort found in Daniel McFadden (1982), G. S. Maddala (1983), and elsewhere, for the analysis of data with limited dependent variables. For exam-

ple, in the treatment where subjects have a single indivisible unit of the private good, they face a simple binary decision. We then model the statistical structure by assuming that utility functions have both uncontrolled fixed components (other than the monetary payoff) that we estimate, and an independent Normally distributed random component. Consistent with terminology elsewhere, we call the fixed components the *altruism* and *warm-glow* effects, which we differentiate below.

The *altruism* effect measures the additional utility a subject gains from increasing the monetary payoff to other subjects by one unit (Ledyard, 1995). Formally, an altruist's utility is modeled as a convex combination of the group payoff and his private payoff. The *warm-glow* effect measures the additional utility a subject gains from just the act of contributing a unit of his endowment (James Andreoni, 1988). Altruistic behavior is present in our data if contributions increase with the public good value, other factors held constant. Warm-glow effects are present if contributions increase with an increase in the difference between the public good value and the token value, other factors held constant. Because we separately vary both the token values for individuals and the public good values, we can identify the effects on contribution rates of these two components of the utility function. This is described in detail in Section II, subsection C.

A. Some Baselines

We first present three different baseline error rates. This gives a rough calibration of a lower bound on the amount of *noise* in the

TABLE 2—THE FREQUENCY OF SPLITTING WHEN THE
ENDOWMENT IS NINE AND $DIFF > 0$.

	Early	Late
Inexperienced	0.36	0.19
	(182)	(176)
Experienced	0.21	0.07
	(180)	(170)

experiment. By noise, we mean the percent of observed decisions that appear incongruous with nearly any currently accepted theory of rational decision-making. We also compare our baseline with baselines observed elsewhere.

Splitting. By splitting, we mean that a subject contributes some fraction of his or her endowment, but not all of it. This is only a possibility in half of our data, the data where subjects have a divisible endowment. Because of the linear structure of the environment, such behavior is not rational even if a subject is altruistic or experiences an additive warm-glow effect. A subject who plays optimally in this environment will always contribute either all or none of his or her endowment, the choice depending on $r_i - V$.[5]

Table 2 shows the frequency of splitting in the experiments where subjects could split. One can see two striking features: first, splitting is more prominent among inexperienced subjects and in the early periods of each ten-period game; second, splitting almost never occurs when subjects have $r_i - V < 0$. Most splitting can be accounted for by inexperienced subjects who have a dominant strategy to free ride.[6]

---

[5] There are possible rationalizations for splitting that we do not consider here. Kay-Yut Chen (1994) constructs a model in which subjects do not know the payoff they will get from their contribution decisions until they have made their choice. In that case, splitting serves a diversification role. It may also be possible to rationalize splitting if the warm-glow (or altruism) effect is nonlinear in contributions.

[6] Splitting is heavily concentrated among a few subjects. Only three of the subjects account for 30 percent of all observations of splitting, and six of the subjects account for over half of such observations. At the other end of the scale, nearly 40 percent of the subjects either never split or split only one time (out of 40 chances).

These findings contrast somewhat with those of Isaac et al. (1984), who observe splitting in well over half the decisions in their data. Furthermore, in some of their experiments the frequency of splitting does not decline over the course of the ten periods. (See Palfrey and Prisbrey [1993] for details.)

Spite. If cooperative behavior (altruism, warm glow, or reputation building) is the main driving force behind the past findings of over-contribution, then we should not observe free-riding from subjects with $r_i - V < 0$. To the extent that violations of dominant strategies to contribute are observed, they might be attributed to effectively random behavior.[7] This gives us a second kind of baseline, called spite (Saijo and Nakamura, 1995). In our experiments, 4 percent of the decisions violate the dominant strategy to contribute when $r_i - V < 0$. This number is quite stable across periods and across the experience treatment.

Sacrifice. In one treatment, $V = 3$, the group optimum does not always occur when everyone contributes. In particular, the group payoff is maximized when subjects contribute if and only if $r_i \leq 4V = 12$. A subject who contributes when $r_i - 12 < 0$ sacrifices more than the entire group benefits. It is hard to imagine any circumstances in which such behavior can be rationalized, except, perhaps, if the warm-glow effects from contributing far outweigh private incentives. Surely such behavior cannot be rationalized for altruists, whose utility is a convex combination of group benefits and private benefits. The frequency of this type of contribution also provides, in a slightly different way, a lower bound on the amount of noise. Among inexperienced subjects, sacrifice occurs with the same frequency as spite, but virtually disappears with experience (1 observation out of 129).

In summary, the kind of behavior that cannot be explained easily with simple models of warm glow or altruism occurs only rarely in our data, and mostly disappears with experience.

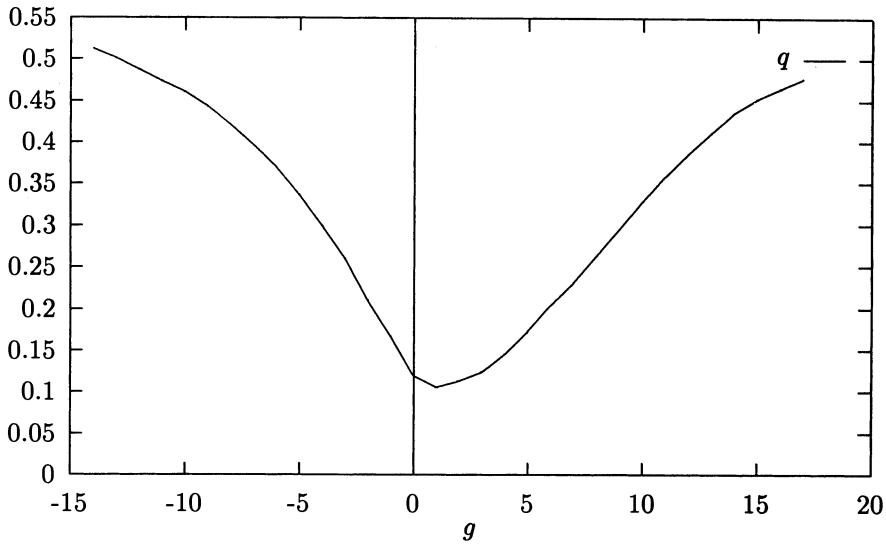[7] However, as we show, some of this may be attributable to a negative warm-glow effect in some individuals.

FIGURE 1. CUTPOINT ANALYSIS: FREQUENCY OF DEVIATIONS FROM THE g-OPTIMAL DECISION RULE

Notes: For each hypothetical warm-glow effect, g, the graph shows the frequency of deviations from the g-optimal decision rule, q. The value $g = 1$ has the lowest associated q.

B. A Simple Model

For a first look at the data, consider the following very simple model of behavior. Assume that all subjects are identical and that they contribute if and only if the difference between their token value and the public good value is less than or equal to some critical value, or cutpoint, g, but that they randomly deviate from this decision rule some fraction of the time, q. Call g the warm-glow effect; e.g., if $g > 0$, then the interpretation is that a subject gains g solely from the act of contribution. Given a fixed value of g, a subject's g-optimal decision rule is:

$$R = \begin{cases} contribute & \text{if } (r_i - V) < g \\ keep & \text{if } (r_i - V) > g \\ keep\,or\,contribute & \text{if } (r_i - V) = g. \end{cases}$$

Despite its simplicity, this class of (g, q) models encompasses a variety of behavior, from completely random decisions ($q = 1$) to the standard model of completely selfish behavior with no error at all ($g = 0$, $q = 0$). From our data, we can estimate the maximum-likelihood values of (g, q) simply by finding

that value of g for which the observed frequency of deviations from the g-optimal decision rule is minimized. Within this very simple class of models such a value of g best describes the data. Figure 1 graphs the observed frequency of deviation from the g-optimal decision rule, for each integer value of g in the range between -15 and 20. The best estimate is $g = 1$, at which the deviation rate is $q = 0.11$. The standard "selfish" model, $g = 0$, is nearly as good, with a deviation rate of $q = 0.12$.[8] The implication of this very simple analysis is that an aggregate warm-glow effect exists, but it is small in magnitude.[9]

There is overcontribution relative to the selfish theory, but much, if not all, of this overcontribution seems to be explainable as

---

[8] Even though the difference in the deviation rate is small, a likelihood ratio test rejects the $g = 0$ model in favor of the $g = 1$ model. The χ^2 statistic is 107.47 with 1 degree of freedom and $n = 2,560$.

[9] The dollar equivalent of the difference between $g = 1$ and $g = 0$ is one cent, in the sense that $g = 1$ corresponds, in experimental payoffs, to behavior in which a subject is willing to contribute his or her endowment if and only if the value of the endowment exceeds the value of the public good by no more than one cent.

noise rather than some systematic component of the decision rule. In the next sections, we examine the nature of the decision rule in detail, giving more consideration to the structure of errors generating deviations, to possible heterogeneity across individuals, and to the role of other factors such as experience and altruism, that are likely to affect contribution decisions.

C. The Probit Model

The probit model provides a standard way to measure the probability of contribution as a function of the different treatment variables, such as the individually assigned token values, the public good value, and experience. The structural model underlying this analysis is the following. We assume that the utility player i gets in period t from contributing x_{it} units of the private good is:

$$U_i(x_{it}, x_{-it})$$

$$= V_t \sum_{j=1}^{N} x_{jt} + (g_i - r_{it})x_{it} + r_{it}w_{it}$$

$$+ \alpha_i \left[(N-1)V_t \sum_{j=1}^{N} x_{jt} \right.$$

$$\left. + \sum_{j \neq i} [(g_j - r_{jt})x_{jt} + r_{jt}w_{jt}] \right],$$

where
  V_t is the public good value in period t,
  g_i is player i's warm-glow term,
  r_{it} is player i's token value in period t,
  w_{it} is player i's endowment of tokens in period t,
  α_i is player i's altruism term, and
  N is number of players in i's group.

Finally, in order to estimate the model we assume that for each of subject i's decisions at period t there is a random component, ε_{it}, that is added to the warm-glow term. This error term represents some random added propensity for the subject to either contribute or not contribute. We assume that the ε_{it}'s are independent, identical, Normally distributed random variables with mean zero and

variance σ^2. A subject contributes if and only if

$$\varepsilon_{it} \geq (r_{it} - V_t) - g_i - \alpha_i(N-1)V_t,$$

where the right-hand side contains all the elements of the subject's utility function that determine his or her choice x_{it}.

Accordingly, we estimate a probit model, where the probability of contributing a unit of the endowment is given by the cumulative Normal transformation of a linear function of the independent variables in the model. Given our specification of the decision rule of the subject, our independent variables are:
  • a constant term, which we call *constant*;
  • the difference $(r_i - V)$, which we call *diff*; and
  • the value of the public good, V.

In addition, we include three other variables that were controlled in the experiment:
  • *exper*, for experience, which takes on a value of zero for decisions in the first ten-period sequence with a given public good value, and one for decisions in the second ten-period sequence of the same public good value;
  • *endow*, which takes on a value of zero if the endowment is indivisible and one if it is divisible; and
  • *period*, which takes on values from one to ten.

D. The Representative Subject Model

We present estimates from two probit models which differ only in which independent variables are included. Note that in these representative subject models, the warm-glow and altruism effects are implicitly assumed to be the same across individuals. An observation is a contribution decision in a single period.[10]

---

[10] We pool observations across all experiments. Decisions in the divisible endowment treatment (*endow* = 1) are coded as either 0 or 1, depending on whether subjects contributed less than half or more than half their endowment of tokens in a given period, respectively. Similar conclusions obtain when the two endowment treatment samples are estimated separately. This is addressed in detail in the next section, where some minor differences are also discussed.

TABLE 3—ESTIMATES FROM PROBIT MODELS

	Probit models		
	1	2	3
diff	−0.25	−0.17	−0.27
	(−27.63)	(−8.12)	(−9.60)
exper.d		−0.059	−0.077
		(−3.23)	(−3.45)
endow.d		−0.034	0.017
		(−1.91)	(0.66)
period.d		−0.0079	−0.0096
		(−2.53)	(−2.51)
constant	0.55	0.52	See Figures
	(6.31)	(3.94)	2 and 3
exper		0.010	0.025
		(0.12)	(0.26)
endow		−0.046	
		(−0.55)	
V	0.0077	0.0089	−0.0020
	(0.90)	(1.02)	(−0.19)
period		0.0070	0.0058
		(0.46)	(0.34)
log likelihood	−810.23	−796.87	−588.92
observations	2,560	2,560	2,560
percent correctly predicted	86.45	86.60	91.48

Notes: In each probit model the dependent variable is the binary investment decision variable. Under each coefficient is the asymptotic *t*-statistic. Variables appended with *.d* are interactions with *diff*. Probit models 1 and 2 assume identical fixed effects for all individuals (homogeneity). Probit model 3 estimates separate individual fixed effects for each of the 64 subjects (heterogeneity). These individual effects are displayed in Figures 2 and 3.

The first column of Table 3 reports the results of estimating the probit equation including only the variables *constant*, *diff*, and *V*. Given the specification of the individual utility functions, the coefficient of *constant* is an estimate of the warm-glow effect divided by the standard deviation of the error term, or g/σ. The coefficients of *diff* and *V* are estimates of $-1/\sigma$ and $\alpha(N-1)/\sigma$. Thus, through algebraic manipulation, we can directly obtain an estimate of the warm-glow effect, g, and the altruism effect, α.

The results are clear. Both estimates have the predicted positive sign, but the coefficient of *V* is so small that the altruism parameter is not significantly different from zero. The coefficients of *constant* and *diff* are both highly significant, indicating a significantly positive

warm-glow effect with g approximately equal to 2.21.[11]

The estimate of the warm-glow term can be interpreted in the following way. Define a cutpoint as the difference between the token value of a subject and the public good value at which our prediction of subject behavior switches from noncontribution to contribution, given specific values of the other independent variables in the model.[12] Such a computation gives a cutpoint of approximately 2.5 token value units if $V = 10$. In other words, on average, with all other variables held fixed at these levels, subjects can be expected to contribute half their endowment when $diff = -2.5$.

It is instructive to contrast this estimate to the one in the previous section, based on the very simple, two-parameter (g, q) model. With that model the probability of contribution is $1 - q$ if $diff > g$ and equals q if $diff < g$. Since the cutpoints estimated under the two models differ, i.e., 1 for the (g, q) model versus 2.5 for the probit model, an obvious question is: which model is better?

This is also a question that is relevant to other recent efforts to estimate models of subject decision errors in experiments. One class of models that has been explored is of the (g, q) variety. These are called constant error models because the probability of a decision error is assumed to be independent of other variables in the model.[13] Another class of models, one that includes the probit model, assumes that decision errors occur more frequently when subjects are nearly indifferent between choices.[14]

---

[11] In fact, the *t*-statistic for g depends on the variances of both the coefficients *constant* and *diff*, and the *t*-statistic for α depends on the variances of both the coefficients *diff* and *pval*. These can be obtained using Taylor series approximations as explained in Jan Kmenta (1971 p. 444). The resulting *t*-statistic for g is 5.6010, and the *t*-statistic for α is 0.9097.

[12] That is, the estimated cutpoint will depend on V in this model.

[13] See, for example, Richard D. McKelvey and Palfrey (1992), Richard T. Boylan and Mahmoud A. El-Gamal (1993), and David W. Harless and Colin F. Camerer (1994).

[14] For example, quantal response equilibrium as defined by McKelvey and Palfrey (1995). Notice that the probit model we propose to explain the data is formally equivalent to a probit response specification of quantal response equilibrium.

Here we see that the estimated warm-glow term is more than twice as large in magnitude in the probit model compared with the constant error (g, q) model. Which estimate is better? There are several ways to conduct such a test, and in all those that we tried, a likelihood ratio test shows the probit model to be the clear winner, at highly significant levels. For example, we conducted a likelihood ratio test between the probit model including only the constant and *diff* variables and the (g, q) model with $g = 1$ and $q = 0.105$. To give the (g, q) model the benefit of the doubt, we assign the likelihood of contribution at the cutpoint $(diff = 1)$ to simply equal the empirical frequency. The likelihood ratio is equal to 93.36. Since the two models are strictly non-nested, we use the Quang H. Vuong (1989) adjustment (rather than a standard chi-square test) to conduct a formal statistical test. This produces a z-statistic of 7.30 (significant at $p < 10^{-12}$).

We next run a probit including the additional control variables *exper*, *endow*, and *period*, and also including the interaction of these variables with *diff*.[15] (See column 2 of Table 3.) The interaction coefficients measure the effect of the variables on the coefficient of *diff*, with a negative coefficient indicating that the variance of the random utility term is getting smaller. Behaviorally, this lower variance translates into more predictable behavior by subjects, or steeper probit response curves.

Not surprisingly, the interaction coefficients for both the experience variable and the period variable show such an effect, indicating that subject behavior is becoming more predictable over time. Also of interest is the fact that none of the noninteraction coefficients are significant. Jointly, this implies that the overall effect of experience and repetition is to reduce aggregate contributions, but that this reduction effect is *indirect* and due to the combination of reduced variance and the fact that the warm-glow level is positive. The estimated difference between cutpoints for inexperienced subjects in round one and experienced subjects in round ten is quite large, with the estimated warm-glow term dropping by nearly

50 percent from 2.7 to 1.6. This suggests that subject confusion may indeed account for a large portion of the contributions by inexperienced subjects.[16]

The bottom lines from the aggregate probit analysis are: (1) there is strong evidence for a warm-glow effect leading to voluntary contribution, and (2) there is no significant evidence for an altruism effect. The results also show that much of the decline in contribution from experience and repetition is due to decline in error rates rather than a change in the underlying decision rule. As such, the decline in error rates is a possible explanation for the decay effects observed in some past experiments, an explanation that avoids any recourse to models of reputation building or repeated games.[17]

E. The Heterogeneous Subjects Model

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 many other economics and decision experiments (McKelvey and Palfrey, 1992; El-Gamal and David M. Grether, 1995) and in public goods experiments (Isaac et al., 1984).

Here, the aggregate analysis of the previous section is broken down at the individual level by including a dummy variable for each individual, from which we can estimate the actual distribution of individual warm-glow effects.[18] The last column of Table 3 reports the coefficients for the included variables,

---

[15] Interactions with V are not included because the effect of V is insignificant.

[16] The coefficient on the *endow* treatment variable is insignificant and the coefficient on the interaction between *endow* and *diff* is very small (<.01) and barely significant at the 5-percent level. In the later analysis with individual effects, this small effect vanishes.

[17] This also provides a possible explanation for Andreoni's (1988) finding that in a random matching design, there is less decay than in the standard repeated-group design. This could happen if subject learning occurs more slowly in the random matching design, which is plausible since the random matching protocol introduces another source of noise in the feedback received by subjects after each period of play. See Palfrey and Prisbrey (1996) for additional evidence for this explanation.

[18] There are other conceivable sources of heterogeneity in these experiments, including cohort effects, nonlinear warm-glow terms, different varieties of altruistic preferences, or differential error rates across subjects, but an exploration of multidimensional heterogeneity is well
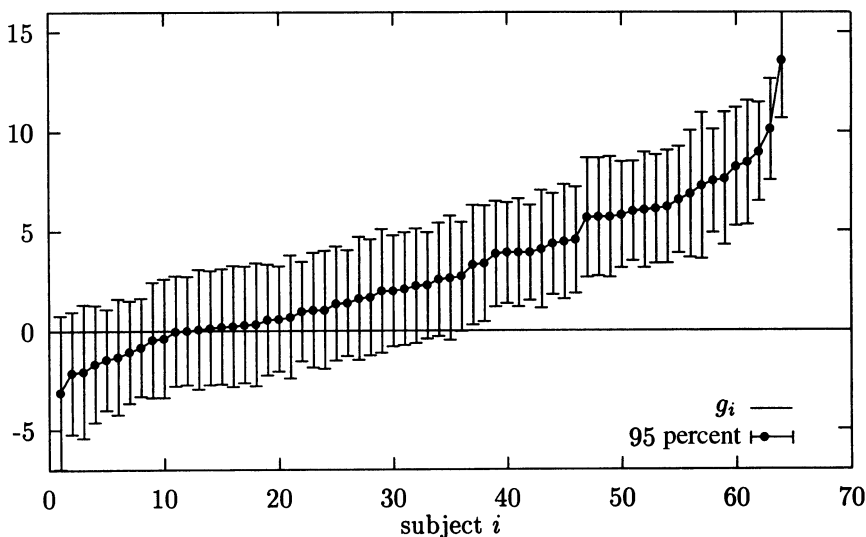
FIGURE 2. INDIVIDUAL WARM-GLOW EFFECTS: PERIOD 1, INEXPERIENCED

Note: The estimated individual warm-glow effects, g_i, for our 64 subjects (inexperienced/period 1).

excluding the coefficients for the 64 individual warm-glow effects. A likelihood ratio test shows clearly that the individual effects are statistically significant at any reasonable level of significance. The χ^2 statistic is 415.9 with 63 degrees of freedom. Thus, we reject the representative subject model in favor of the heterogeneous subject model. The information from the individual coefficients is summarized in Figures 2 and 3, which graph g_i with 95-percent confidence intervals, for inexperienced and experienced subjects, respectively. Each individual cutpoint is calculated from the probit coefficients, in a manner similar to the computation of the aggregate cutpoint in the previous section.[19] The

beyond the scope of this paper. For example, simple classification analysis in Palfrey and Prisbrey (1993) suggests the possibility of differential error rates, which we have chosen not to model explicitly here. Nevertheless, we are confident that our specification captures the key component of subject heterogeneity in these experiments. As with any estimation, our results are subject to the usual caveat about other additional (unmeasured) sources of heterogeneity.

[19] The confidence intervals were derived using an estimate of the variance of g_i. The estimate was created using a Taylor series approximation as described in Kmenta (1971 p. 444).

median warm-glow effect is 2.3 for inexperienced subjects in period one and 1.4 for experienced subjects in period ten, which is very close to the aggregate results of the previous section. Considerably less than half the subjects have a warm-glow term that is significantly greater than zero. No subject has one that is significantly negative.

The distribution of cutpoints in the experienced, period-ten treatment is clearly less dispersed and has a lower median than the inexperienced distribution, which simply reflects the significant effect of those variables on reducing error rates, as discussed earlier. The decisions are moving in the direction of the predictions of the selfish model, where the warm-glow effect is zero. The confidence interval around each individual cutpoint is bigger because of the compounded effect of the variance of *period.d.*

The *endow* variable was excluded because otherwise the model is not identified, i.e., *endow*, *pval*, and the individual dummies are collinear. To test for any effects due to the endowment, we separately estimate model 3 for the two subsamples defined by the *endow* = 0 (binary endowment) and *endow* = 1 (divisible endowment) treatments. The results are reported in Table 4, and the estimated in-
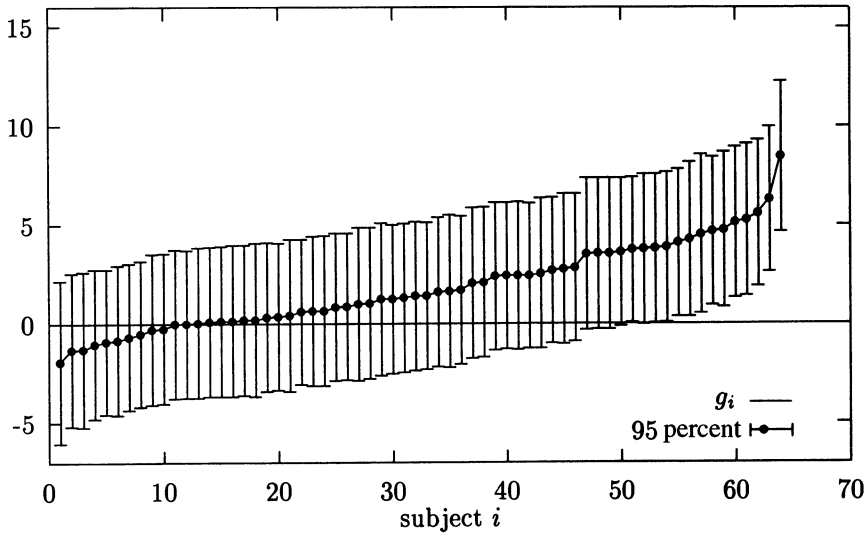
FIGURE 3. INDIVIDUAL WARM-GLOW EFFECTS: PERIOD 10, EXPERIENCED

Note: The estimated individual warm-glow effects, g_i, for our 64 subjects (experienced/period 10).

dividual warm-glow terms are reported in Figure 4.[20]

The first column of Table 4 is the same as the last column of Table 3. The second column gives the parameter estimates for the binary endowment treatment, and the last column gives the parameter estimates for the divisible endowment treatment. The similarities and differences are as follows.

First, the three key findings of this heterogeneous probit analysis are the same in the two separate samples: (a) *diff* is highly significant in both treatments, and of the same order of magnitude—hence warm glow is significant and of the same importance in both samples; (b) *V* is not significant in either treatment, hence altruism is insignificant in both samples; and (c) the implied distribution of individual warm-glow terms is the same as it was in the pooled estimation.[21] These similarities are not surprising, given the low splitting rates observed in the experiment, and the concentration of these splitting rates in a small subsample of the population.

There is one difference between the separately estimated models, one which is of relatively minor consequence. The interaction term between *diff* and *exper* is significant for the divisible token sample, but not for the other sample. The coefficient is also larger in magnitude. This suggests that error rates in the divisible endowment condition may start slightly higher and decline faster with experience. We speculate that this is either due to natural sample variation (recall that there are only 32 subjects in each treatment), or indicative of a small treatment effect. The subjects in the divisible endowment treatment have more choices (ten instead of two) and receive finer feedback. The additional choices may be more confusing initially, while the feedback may enable subjects to gain experience faster.

---

[20] The individual warm-glow terms are evaluated at the means of the other independent variables.

[21] A Kolmogoroff-Smirnoff test on the equality of the two distributions of individual warm-glow terms, one obtained from the pooled sample and the other from separate estimations based on the endowment, indicates that the distributions are not statistically different. The test statistic is 0.0625. The critical value for $\alpha = 0.05$ is approximately 0.2404. We also estimated the *endow* = 1 data using an ordered probit model, and found no significant differences between those estimates and the estimates in Table 4 and Figure 4.

TABLE 4—HETEROGENEOUS PROBIT
SUBSAMPLE ESTIMATES

	Pooled	Binary only	Divisible only
diff	−0.27	−0.31	−0.23
	(−9.60)	(−8.47)	(−7.24)
exper.d	−0.077	0.003	−0.22
	(−3.45)	(0.10)	(−5.13)
endow.d	0.017		
	(0.66)		
period.d	−0.0096	.0077	−0.011
	(−2.51)	(−1.43)	(−2.06)
constant		See Figure 4	
exper	0.025	−0.059	0.12
	(0.26)	(−0.42)	(0.85)
V	−0.0020	0.015	−0.015
	(−0.19)	(0.95)	(−1.06)
period	0.0058	0.015	−0.0089
	(0.34)	(0.60)	(−0.37)
log likelihood	−588.92	−287.24	−288.63
observations	2,560	1,280	1,280
percent correctly predicted	91.48	91.72	91.17

Notes: In each probit model the dependent variable is the binary investment decision variable. The column two estimates are computed using data from the treatment where subjects were endowed with one token. The last column uses data where subjects were endowed with nine tokens. Under each coefficient is the asymptotic *t*-statistic. Variables appended with *.d* are interactions with *diff*. The individual fixed effects for each of the 64 subjects (heterogeneity) are displayed in Figure 4.

The significant coefficient may also reflect the fact that splitting declines sharply with experience.

F. *The Effect of Noise*

The individual analysis described above also allows a comparison of the relative importance of noise to the warm-glow and altruism effects on the subject's ultimate decision. Past experimental designs have been unable to differentiate these effects. In past designs, subjects were only observed making decisions when they had a dominant incentive to keep. Any error that was made was necessarily a contribution, and the experimenter had no way to differentiate the noise from contributions due to a warm glow or altruism. Our design and the properties of the probit model give us a way to control for and measure the magnitude of the noise.

Take our estimated distribution of individual cutpoints, from the pooled estimation, as fixed. Then, from this calculate a predicted frequency of contribution if subjects were assumed never to make errors relative to their cutpoint. For example, subjects with a cutpoint of four are predicted to contribute their entire endowment if and only if $r_i - V \leq 4$. Since the estimated distribution of cutpoints varies across periods and experience level, we can generate profiles of aggregate contribution rates as a function of *diff* for each period and each experience level. The *no-noise* curves in Figures 5 and 6 display these profiles for the two polar extremes, respectively, *period one/inexperienced* and *period ten/experienced*. Other profiles, which include the effect of errors on the frequency of contribution, can be computed directly from the probit scores of each individual at each value of *diff*. The *noise* curves in Figures 5 and 6 display these profiles for identical extremes.

The difference between the no-noise and noise curves represents the amount of contribution that is attributable to noise, for each different value of *diff*. In particular it vividly illustrates the measurement problem inherent in experiments where all subjects are given the same positive values of *diff*. For example, for *diff* = 5, we estimate that more than half the observed contributions of experienced subjects are due to random variation. Of course, a reverse effect occurs at values of *diff* below the average cutpoint (consistent with the observations of Saijo and Nakamura, 1995). However, according to our estimates, this reverse effect seems to be small in magnitude, since most warm-glow fixed effects are positive.

It is also possible to apply these measurements to past experiments that used a fixed value of r_i for all subjects and for all contribution rounds. For example, Isaac et al. (1984) conducted several experiments in which all subjects' marginal rate of substitution between the public good and the private good, r_i/V, equaled 3.33. Because of differences in subject pools, payoff magnitudes, and other design factors, the translation of this marginal rate of substitution into our *diff* parameter in our experiment is admittedly very rough. With this caveat in mind, $r_i/V = 3.33$ corresponds to $V = 3$, $r_i = 10$, or $V = 6$,
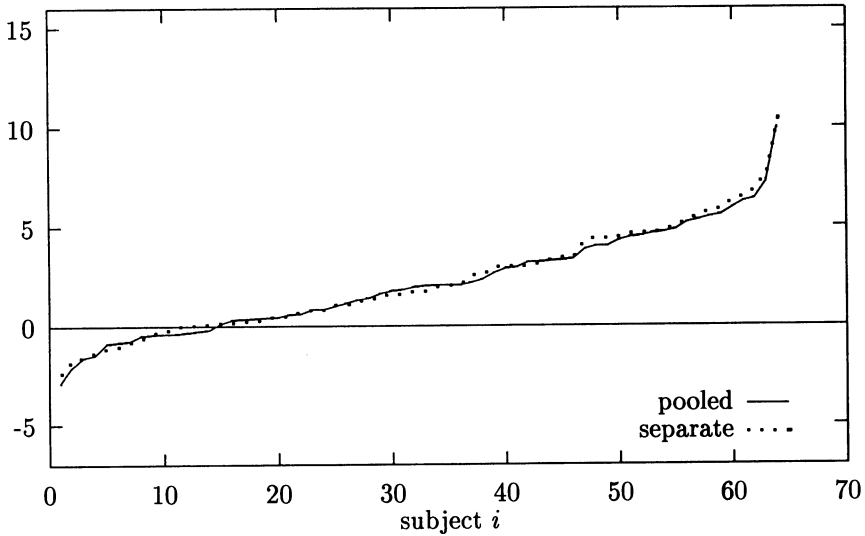
FIGURE 4. INDIVIDUAL WARM-GLOW EFFECTS: POOLED VS. SEPARATE

Note: The estimated individual warm-glow effects from the separate estimations by endowment condition (Table 4 columns 2 and 3) compared to the pooled estimates (Table 4 column 1).

$r_i = 20$, which yields a wide range of predictions, depending which one is used. Using our measurements, the range of predicted contribution lies between 4 percent and 20 percent for inexperienced subjects, well over half of which is attributable to error. Predictions for experienced subjects lie in the range of 10 percent or less.

This estimated contribution rate is somewhat less than than was observed by Isaac et al., at least for inexperienced subjects. We did, however, successfully replicate the Isaac et al. results in an additional experiment using our protocol and a fixed r_i across subjects.[22] This suggests that some of the overcontribution in past experiments may be due to the use of a degenerate distribution of private values.[23] Andreoni (1988) conducted experiments similar to those of Isaac et al. (1984), but with five-person groups and $r_i/V = 2$. Comparisons of our data to his lead to similar conclusions.

These results can also be compared to recent findings by Andreoni (1995) in an experiment which was also motivated by the problem of differentiating errors. His design consisted of two treatments: one where subjects were paid what they earned in the experiment, and another where subjects were paid by a fixed formula based on the rank of their payoff. The latter treatment is assumed to remove much of the altruistic incentive for contribution.[24] Otherwise, the experiment is run in the usual

---

[22] This is not included in the present paper for reasons of space. See Palfrey and Prisbrey (1993) for details. That paper also presented results from a treatment in which subjects were informed of their fellow group members' token values, which was conducted to test whether the incomplete information in the present experiment was responsible for the slightly lower contribution rates. The finding was that this additional information did not increase contribution rates, rejecting the hypothesis that incomplete information leads to more free-riding.

[23] One theoretical explanation for this difference would be that warm-glow effects are subject to diminishing returns and that these effects are cumulative over the course of the experiment. In the nondegenerate design, subjects nearly always contribute when they have low values of r_i, so diminishing warm-glow effects would lead to less contribution than in the degenerate designs where r_i is always greater than V. This suggests possible future experiments.

[24] It is not entirely clear why rank payoffs should necessarily remove the warm-glow incentive for contribution.
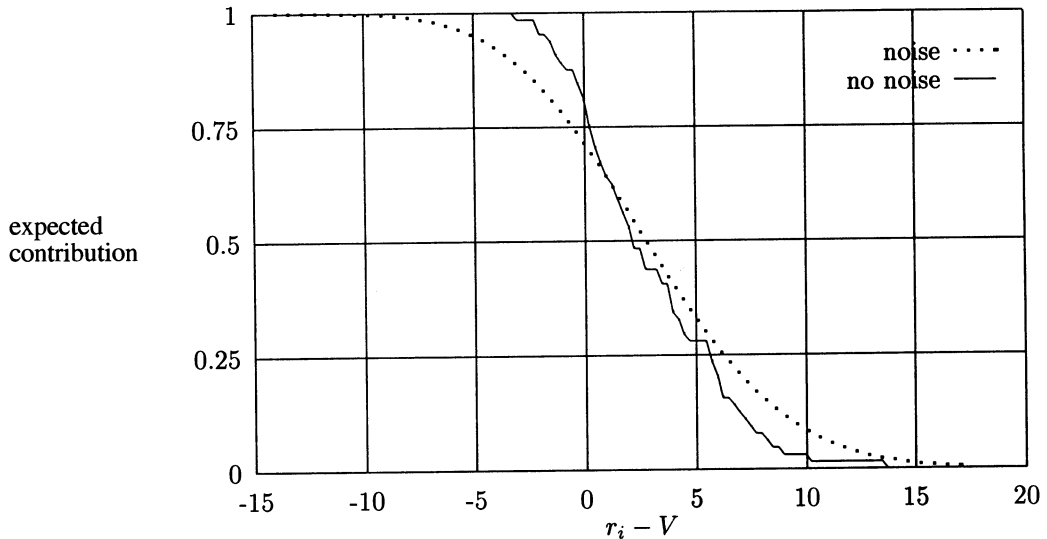
FIGURE 5. EXPECTED SPENDING: PERIOD 1, INEXPERIENCED

Notes: The noise curve assumes subjects make errors as in probit model 3. The no-noise curve assumes the subjects make no errors. The curves here represent an inexperienced population in period 1.

fashion, with homogeneous valuations. He finds that if one attributes all of the contribution in the rank treatment to confusion (what we call noise) this accounts for approximately one-third of the total contribution in the regular treatment.

III. Conclusion

We have designed and carried out an experiment that differs from past public goods experiments in that the marginal value of the private good differs across subjects, and also across periods for the same subject. As a result, individual response functions could be, and were, estimated. That is, we estimated the probability that a particular individual will contribute given his or her value for the private good and the common value for the public good.

In turn, the relative importance of altruism effects, warm-glow effects, and subject error were determined. We found that altruism played little or no role at all in the individual's decision and, on the other hand, warm-glow effects and random error played both important and significant roles. We further measured heterogeneity in the warm-glow effect across

subjects and found that they fell in a fairly wide range of predominantly nonnegative values.

As in past experiments, we found that experience is a significant explanatory variable and leads to declining contribution rates. This decline was shown to be the result mainly of a reduction in the amount of subject decision error combined with a lower variance in the distribution of individual warm-glow effects. It is not due to an overall decline in warm glow. Players do not become significantly more selfish with experience; rather, their preferences as we measure them are relatively stable with respect to experience. Overall, most of the raw data from our experiments corroborate past findings but we offer a much different explanation for these data. The consistent observations include the significance of decay and experience, and the generally very strong responsiveness of contribution rates to the opportunity cost of contribution. There are some differences in magnitude, which we view as minor. The fact that data from past experiments exhibit even higher contribution rates than we observed may be due to the fixed nature of the subjects' valuations in those experiments.
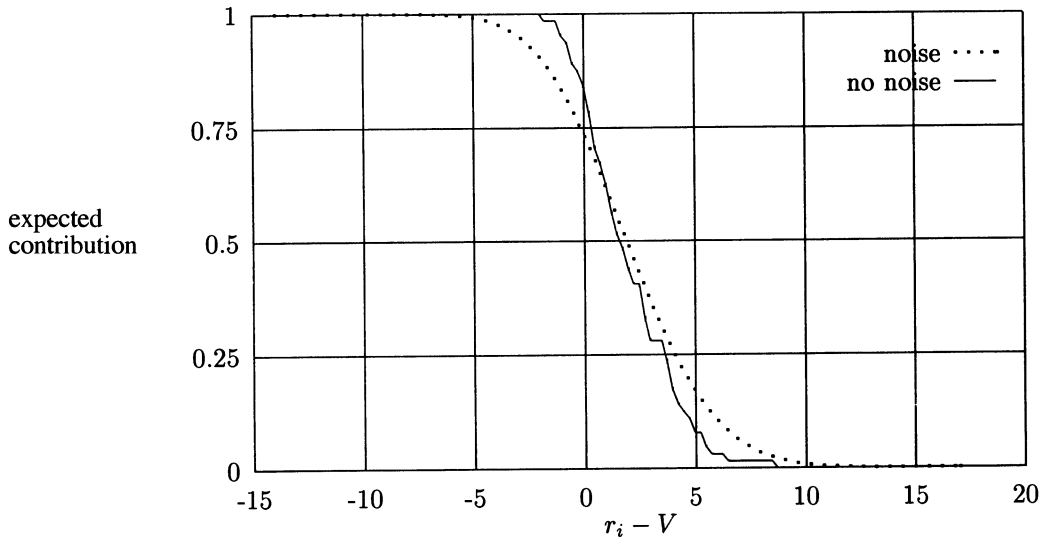
FIGURE 6. EXPECTED SPENDING: PERIOD 10, EXPERIENCED

Notes: The noise curve assumes subjects make errors as in probit model 3. The no-noise curve assumes the subjects make no errors. The curves here represent an experienced population in period 10.

While more research obviously remains to be done before we have a complete picture of the incentives and motivations for individuals to contribute to public goods, quite a bit has been learned here, particularly about the role of nonmonetary components of the utility function and the role of subject decision errors. Besides this, the methodological lesson from this experiment should be clear. It is indeed possible to design experiments where the details of individual decision rules can be accurately measured. And, furthermore, these measurements can be used to distinguish between different theoretical explanations for interesting systematic features in the aggregate data. Given the considerable amount of heterogeneity of behavior across subjects that is known to be characteristic of these experiments, improved measurement at the individual level would seem to be a necessary ingredient to reaching a better understanding of these phenomena.

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 ten rounds. At the end of the last experiment, you will be paid the total amount you have accumulated during the course of all four 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 four experiments.

In each experiment you will be divided into four groups of four persons each. Those groups will stay the same for all ten rounds of the experiment. After each ten-round

experiment, everyone will be regrouped into four 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 four experiments tonight, you will be assigned four different groupings, one for each ten-round experiment.

Rules for Experiment 1

Each round of the experiment you will have nine 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 ten 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 ten rounds. Following each round you will begin with nine 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 AND EXPLAIN HOW TO READ IT]

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 four tokens and

your token value was 12. You would earn 12 + 39 + 60 = 111 FRANCS. If you had spent three tokens you would have earned 9 +

39 + 72 = 120 FRANCS. If you had spent five 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 his or her ID#. In round 2 have each subject KEEP the number of tokens equal to the last digit of his or her ID#. Go over screen display and history display. Tell subjects to refrain from pressing keys for no reason.]
[Keep screen display on.]
[Hand out quiz.]
[Correct quiz answers and read correct answers 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 his or her 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	$(N*15) + (9*\text{your token value})$
1	N tokens	$15 + (N*15) + (8*\text{your token value})$
2	N tokens	$30 + (N*15) + (7*\text{your token value})$
3	N tokens	$45 + (N*15) + (6*\text{your token value})$
4	N tokens	$60 + (N*15) + (5*\text{your token value})$
5	N tokens	$75 + (N*15) + (4*\text{your token value})$
6	N tokens	$90 + (N*15) + (3*\text{your token value})$
7	N tokens	$105 + (N*15) + (2*\text{your token value})$
8	N tokens	$120 + (N*15) + \text{your token value}$
9	N tokens	$135 + (N*15)$

Example:
Suppose everyone else in your group spends 13 tokens in all and you spend four tokens and your token value was 12. You would earn 60 + 195 + 60 = 315 FRANCS. If you had spent three tokens you would have earned 45 + 195 + 72 = 312 FRANCS. If you had spent five 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.]
[Pay subjects in private in separate room and dismiss them one at a time.]

REFERENCES

Andreoni, James. "Why Free Ride? Strategies and Learning in Public Goods Experiments." *Journal of Public Economics*, December 1988, *37*(3), pp. 291–304.

———. "Cooperation in Public Goods Experiments: Kindness or Confusion?" *American Economic Review*, September 1995, *85*(4), pp. 891–904.

Boylan, Richard T. and El-Gamal, Mahmoud A. "Fictitious Play: A Statistical Study of Multiple Economic Experiments." *Games and Economic Behavior*, April 1993, *5*(2), pp. 205–22.

Brookshire, D. S.; Coursey, D. L. and Redington, D. B. "Special Interests and the Voluntary Provision of Public Goods."

Working paper, University of Wyoming, 1989.

Chen, Kay-Yut. "The Strategic Behavior of Rational Novices." Ph.D. dissertation, California Institute of Technology, 1994.

El-Gamal, Mahmoud A. and Grether, David. "Are People Bayesian? Uncovering Behavioral Strategies." *Journal of the American Statistical Association; Applications and Case Studies*, December 1995, *90*(432), pp. 1137–45.

Fisher, Joseph; Isaac, R. Mark; Schatzberg, Jeffrey W. and Walker, James M. "Heterogeneous Demand for Public Goods: Effects on the Voluntary Contribution Mechanism." *Public Choice*, December 1995, *85*(3–4), pp. 249–66.

Harless, David W. and Camerer, Colin F. "The Predictive Utility of Generalized Expected Utility Theories." *Econometrica*, November 1994, *62*(6), pp. 1251–89.

Isaac, R. Mark; McCue, Kenneth F. and Plott, Charles R. "Public Goods Provision in an Experimental Environment." *Journal of Public Economics*, February 1985, *26*(1), pp. 51–74.

Isaac, R. Mark and Walker, James M. "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism." *Quarterly Journal of Economics*, February 1988, *103*(1), pp. 179–99.

———. "Complete Information and the Provision of Public Goods." Discussion Paper 89-18, University of Arizona, 1989.

Isaac, R. Mark; Walker, James M. and Thomas, Susan H. "Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations." *Public Choice*, 1984, *43*(2), pp. 113–49.

Isaac, R. Mark; Walker, James M. and Williams, Arlington W. "Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups." *Journal of Public Economics*, May 1994, *54*(1), pp. 1–36.

Kmenta, Jan. *Elements of econometrics*. New York: Macmillan, 1971.

Ledyard, John O. "Public Goods: A Survey of Experimental Research," in John Kagel and Alvin Roth, eds., *Handbook of experimental economics*. Princeton, NJ: Princeton University Press, 1995, pp. 111–94.

Maddala, G. S. *Limited-dependent and qualitative variables in economics*. Cambridge: Cambridge University Press, 1983.

Marwell, Gerald and Ames, Ruth E. "Experiments on the Provision of Public Goods, II: Provision Points, Stakes, Experience and the Free Rider Problem." *American Journal of Sociology*, January 1980, *85*(4), pp. 926–37.

McFadden, Daniel. "Econometric Models of Probabilistic Choice," in C. Manski and D. McFadden, eds., *Structural analysis of discrete data with econometric applications*. Cambridge, MA: MIT Press, 1982, pp. 198–272.

McKelvey, Richard and Palfrey, Thomas R. "An Experimental Study of the Centipede Game." *Econometrica*, July 1992, *60*(4), pp. 803–36.

———. "Quantal Response Equilibrium for Normal Form Games." *Games and Economic Behavior*, July 1995, *10*(1), pp. 6–38.

Palfrey, Thomas R. and Prisbrey, Jeffrey E. "Anomalous Behavior in Linear Public Goods Experiments: How Much and Why?" Social Science Working Paper No. 833, California Institute of Technology, 1993.

———. "Altruism, Reputation, and Noise in Linear Public Goods Experiments." *Journal of Public Economics*, September 1996, *61*(3), pp. 409–27.

Palfrey, Thomas R. and Rosenthal, Howard. "Testing Game-Theoretic Models of Free Riding: New Evidence on Probability Bias and Learning," in T. Palfrey, ed., *Laboratory research in political economy*. Ann Arbor, MI: University of Michigan Press, 1991, pp. 239–68.

Saijo, Tatsuyoshi and Nakamura, Hideki. "The 'Spite' Dilemma in Voluntary Contributions Mechanism Experiments." *Journal of Conflict Resolution*, September 1995, *39*(3), pp. 535–60.

Vuong, Quang H. "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses." *Econometrica*, March 1989, *57*(2), pp. 307–33.