Can Public Goods Experiments Inform Environmental Policies?

Water Policy Working Paper #2005-017

Prepared by:

Paul Ferraro and Christian A. Vossler*

June, 2005

* Paul Ferraro is affiliated with Georgia State University and Christian A. Vossler is affiliated with the University of Tennessee. Research supported by the Georgia State University Office of Sponsored Programs. Thanks to Krawee Ackaramongkolrotn for software programming, Kwaw Andam for data entry, and Catherine Eckel, Susan Laury, Ragan Petrie and participants at the 74th Annual Meeting of Southern Economic Association and the 2004 North American Regional Meeting of the Economic Science Association for useful comments. The authors gratefully acknowledge financial support for this work received from Georgia Soil and Water Conservation Commission, Contract No. 480-05-GSU1001; and the U.S. Department of Agriculture, Award Document No. 2003-38869-02007-1.

CAN PUBLIC GOODS EXPERIMENTS INFORM ENVIRONMENTAL POLICIES?

Abstract

Understanding behavior in experimental public goods games is fundamental to the work of environmental, behavioral, institutional and policy-oriented economists. Although much research has been devoted to explaining the dynamics of such experiments, the conclusions drawn to date are contradictory. Through the use of a novel experimental design, a theoretical model of behavior, and appropriate econometric methods, we address weaknesses in the current literature and resolve much of the conflicting claims about motives in public goods experiments. Our analysis demonstrates that herders and strong reciprocators are the main contributors to the public good, whereas the role of interdependent utility and warm-glow altruism is weak at best. Further, the oft-observed decay in contributions over rounds is driven by the revocation of cooperation by disappointed strong reciprocators coupled with the herding behavior of confused subjects. We find no evidence that confused subjects learn the dominant strategy over time. The data instead imply that a substantial proportion of subjects do not recognize the tension between the privately optimal strategy and the socially optimal strategy. These results offer insights into improving environmental policy, but also suggest that public goods experiments cannot achieve their full potential as long as the way in which they are implemented in the laboratory leaves most subjects unaware of the social dilemma that experimentalists are trying to induce.

Keywords:
Public goods; experiment; strong reciprocation; herding; dynamic modeling

I. Introduction

Much of environmental policy is designed to induce private individuals to contribute the public environmental good when it is in their private interests to avoid such contribution. For example, when individuals consume freshwater they typically ignore the environmental impacts of removing that water from natural systems (whether or not they are aware of such impacts). Efforts to reduce freshwater consumption can thus try to raise the price of water use to reflect the environmental values of freshwater in situ, or they can try to encourage social norms that frown upon "excessive use" of water (thus generating a private cost to anyone who accepts the norm and violates it). Economists often use laboratory experiments to explore how individuals behave under different public goods scenarios and gain insights into how institutions might be better designed to encourage the provision of environmental public goods.

The large and diverse experimental literature on the private provision of public goods is based principally on variants of the Voluntary Contributions Mechanism (VCM) game. In a typical linear VCM experiment, subjects are given an endowment of "tokens" to be divided between a private account and a public account (Isaac, Walker and Thomas 1984). Contributions to the public account yield a return to each of the n individuals in the group, regardless of their contribution level. If the marginal return from contributing a token to the public account is less than the value of a token kept in the private account, but the sum of the marginal returns to the group is greater than the value of a token kept, the individually rational contribution is zero (i.e., the individual free rides) while the social optimum is realized when everyone contributes their entire endowment to the public account.

In one-shot VCM experiments with a dominant strategy of contributing nothing to the public good, subjects contribute at levels far above the theoretically predicted value: on average, 40-60% of endowments.[1] In repeated-round VCM experiments, contributions start in the range of 40-60% but then decay towards zero (ending around 10% of endowments on average). Although the pattern of initially high contributions rates and subsequent decay is generally observed, the cause of these dynamics is controversial (Ledyard 1995, p. 148).

Possible motives underlying contributions include self-interested weak reciprocity ("reciprocal altruism") to promote contributions from other group members, other-regarding

[1] We focus on classic linear public good games that (1) use the Voluntary Contributions Mechanism to elicit contributions and (2) create a dominant strategy to free ride and a socially optimal strategy to contribute the entire endowment. We believe, however, that the inferences drawn are applicable to other public goods games.

preferences such as interdependent utility, warm-glow altruism, and strong reciprocity ("conditional cooperation"), and decision error stemming from the failure to identify the dominant strategy ("confusion"). The decay in contributions could be a result of learning the dominant strategy or a revocation of conditional cooperation.

Experimental economists have used different approaches to discriminate among these explanations. Most approaches rely on clever manipulations of subject payoffs, changes in the rules of the VCM game, or using results from other games to make inferences about behavior in the VCM game. Unfortunately, the reliance on these methods has resulted in incomplete separation of motives and conflicting conclusions about the factors that drive behavior in the VCM game (Andreoni, 1995; Palfrey and Prisbrey, 1997; Houser and Kurzban, 2002; Goeree, Holt and Laury, 2002; Carpenter, 2004; Fischbacher and Gächter, 2004).

Understanding behavior in the VCM game is more than an academic curiosity. The VCM game and its variants (e.g. Provision Point Mechanism game) are the workhorses of experimental research on the private provision of public goods. Public economists use the VCM to test a variety of public economic theories, behavioralists use the VCM to gain insight into the nature of individual preferences in collective action situations, and institutionalists and policy-oriented economists use the VCM to explore how changes in the rules affect collective outcomes. Continued use of the VCM game without a clear understanding of what drives subject behavior is perilous. In fact, the analysis we present implies that a substantial proportion of subjects in VCM experiments do not recognize the tension between the dominant strategy and the socially optimal strategy: they are simply confused "herders" that take group contributions as an indication of optimal contributions levels (they are erroneously classified as "conditional cooperators" in previous studies). Thus, the internal and external validity of the VCM experiment for making inferences to real-world phenomenon is questionable unless such confused subjects can be identified in an experiment.

Our analysis improves upon previous work in three important ways: (1) we develop a novel experimental design that allows one to better discriminate between contributions stemming from other-regarding preferences and those due to confusion, *without changing the fundamental rules of the VCM game*; (2) we develop a behavioral model of individual contributions that complements the experimental design; and (3) we apply appropriate econometric methods to estimate the unknown parameters of the behavioral model and draw inferences.

In the next section, we survey recent research that endeavors to explain behavior in VCM experiments. In Section III, we present the experimental design. In Section IV, we present aggregate results. In Section V, we develop a model of individual behavior and present our econometric analysis. In Section VI, we discuss the results and present further evidence from a post-experiment "focus group" session. We conclude in Section VII.

II. Previous Experiments

In this section, we highlight recent articles that capture the on-going debate over the relative importance of different behavioral motives in the VCM game.[2] Before describing these articles, however, we wish to ensure that the vocabulary we use is clear.

The term "other-regarding behavior" will be used as an umbrella term to characterize three motives for contributions in the VCM game: (1) "inter-dependent utility" (often called "pure altruism"), which describes a situation in which an individual's utility function is a function of his own payoff and the payoffs of his group members; (2) "warm-glow" (often called "impure altruism"; Andreoni, 1990), which describes a situation in which an individual gains utility from the simple act of contributing to a publicly spirited cause; and (3) "strong reciprocity," which is sometimes referred to as "conditional cooperation." As described by Bowles, Fehr and Gintis (2003, p.1-2), "[s]trong reciprocity is a combination of altruistic rewarding, which is a predisposition to reward others for cooperative, norm-abiding behaviors, and altruistic punishment, which is a propensity to sanction others for norm violations. Strong reciprocators bear the cost of rewarding or punishing but gain no individual economic net benefit from their acts. Strong reciprocity thus constitutes a powerful incentive for cooperation even in non-repeated interactions and when reputation gains are absent because strong reciprocators will reward those who cooperate and punish those who defect….Strong reciprocators' contributions are not contingent upon personal reward and their punishing of defectors is based on the other's behavior, not the punisher's expected net gain from punishing."[3]

In contrast, the term "weak reciprocity" refers to acts of reward and punishment only if these acts contribute to the individual's private economic payoff. Such behavior, often called

[2] There are many good articles published on the VCM game. Because of space constraints, we limit our review to a few articles that summarize well the current debate in the literature.

[3] Other authors have described similar behavior without any special terms. For example, Andreoni (1995) hypothesized that some of the decay in contributions over rounds in the VCM game might stem from frustrated subjects who revoke their cooperation when they observe that others are not cooperating.

"reciprocal altruism" in the evolutionary biology literature, comprises strategic behavior aimed at securing private benefits. We include such behaviors as "reputation building" and "strategic donor leadership" under the term weak reciprocity.

The term "confusion" is an umbrella term to characterize behavior that stems from subjects' inability to discern the nature of the game in which they are playing. Such players are unable to discern the dominant strategy in a linear VCM game. We hypothesize that subjects become less confused in the VCM game through two types of learning: (1) "adaptive reinforcement," or "hill climbing," by which subjects search for the profit-maximizing strategy based on information from previous rounds; and (2) "herding," or "imitation," by which subjects simply behave as they perceive most other individuals behave (i.e., "the trend is your friend"). "Herding" refers to an individual's perceived lack of understanding of a situation and subsequent attempt to use information collected and communicated by others as a behavioral compass.

Andreoni (1995) developed a VCM-like game that fixes the pool of payoffs and pays subjects according to their contributions to the public good. The person who contributes the least is paid the most from the fixed pool. Thus contributions to the "public good" in this game do not increase aggregate benefits, but merely cost the contributor and benefit the other group members. Andreoni uses behavior from the ranking games to infer that both other-regarding behavior and confusion are both "equally important" motives in the VCM, although the relative importance of each motive cannot be precisely identified through the experimental design.

Houser and Kurzban (2002) continued Andreoni's work with a clever experimental design that includes: (1) a "human condition," which is the standard VCM game; and (2) a "computer condition," which is similar to a standard VCM game except that each group consists of one human player and three non-human computer players and the human players are aware they are playing with computers. Each round, the aggregate computer contribution to the public good is three-fourths of the average aggregate contribution observed for that round in the human condition. By making the reasonable assumption that other-regarding preferences and confusion are present in the human condition, but only confusion is present in the computer condition, Houser and Kurzban conclude that confusion accounts for about half of all public good contributions in the standard VCM game and that the decay across rounds is almost entirely due to reduced confusion, rather than any declines in other-regarding behavior. The studies by

Andreoni and by Houser and Kurzban do not allow one to precisely discriminate among different kinds of other-regarding preferences or confusion behaviors.

Since the Houser and Kurzban design is closely related to our own, we wish to focus on some aspects of their approach that make interpreting their results difficult. First, as noted by the authors, average contributions in their human condition (n=20 individuals in 5 groups of 4) display an unusual pattern: contributions start at about 62% of endowment and only decline to 51% of endowment by the 10th and final round. With this unusual pattern of decay (contributions typically decline much more) and only ten rounds for evaluating behavior, it is possible that their results may not readily extend to typical VCM experiments.

Second, the identification of "confusion" contributions in their design relies on the assumption that contributions in a given round are independent of the history of group contributions. If they are not, individual subjects are not independent observations and merely presenting all computer condition subjects with three-fourths of the average aggregate contributions from the human condition thwarts important dynamics. Ashley, Ball and Eckel (2003) analyze raw data from previously published VCM experiments and find that the history of group contributions matters quite a lot (see also Carpenter 2004).

Third, and related to the role of the history of contributions, the computer condition changes the standard VCM game beyond simply grouping a human with automata. Human subjects in the computer condition observe their group members aggregate contribution *before* they make their decision in a round (as opposed to after they make their decision, as in the human condition). If the history of contributions affects both confused and other-regarding subjects, then such a change in design can also affect the comparability of the two treatments.[4]

Palfrey and Prisbrey (1997) developed an alternative experimental design that, when combined with a few behavioral assumptions, allows the authors to separate the effects of inter-dependent utility, warm-glow and confusion. Their design changes the standard VCM game by randomly assigning different rates of return from private consumption each round, which enables the measurement of individual contribution rates as a function of that player's investment costs (n=64). A key assumption in their analysis is that other-regarding preferences take only the form

[4] A similar design developed and applied independently to a single-shot VCM game concluded that about half of contributions were a result of confusion (Ferraro et al. 2003), whereas 74% of first-round contributions in Houser and Kurzban were a result of confusion. In our design that follows, we also find that about 50% of the first round contributions were a result of error. The estimates may differ because of the information about computer contributions that confused subjects had in round one of Houser and Kurzban's design.

of interdependent utility and warm-glow altruism, and the strength of both motives does not decline over rounds. The authors conclude that (1) interdependent utility has no detectable effect on behavior and (2) warm-glow is present but small. The decay in contributions is, by assumption, attributed to reductions in confusion.

Goeree, Holt and Laury (2002) use an alternative VCM design in which group size is either two or four and the "internal" return of a subject's contribution to the public good to the subject may differ from the "external" return of the same contribution to the other group member (n=32). The authors estimate a logit choice model of noisy decision-making based on data from a series of one-shot VCM games (no feedback) in which the internal and external returns are varied. In contrast to Palfrey and Prisbrey, they conclude that interdependent utility motivates subject contributions rather than warm-glow altruism. They also find that errors ("noise") play a role in explaining contributions.

Carpenter (2004) moves beyond characterizing confusion as simply statistical noise and develops a theoretical model of herding ("imitation") in the VCM game, which he describes as an attempt "to take advantage of the information acquired and processed by others" (p.396) by "copying the most observed behavior in a population" (p.395). Carpenter's dynamic replicator model of imitation predicts that were subjects (n=165) to observe the individual contributions of their group members rather than the just the aggregate level of contributions, the decay of contributions would be more rapid because herders would have more information and that information would facilitate their move towards zero contributions. Carpenter's experimental results confirm his theoretical prediction.

Fischbacher and Gächter (2004) correctly point out that that previous experimental analyses do not allow for the presence of "conditional cooperators," or, in our terms, strong reciprocators. In Fischbacher and Gächter's "P-experiment", they ask subjects to specify, for each average contribution level of the other group members, how much they would contribute to the public good. The experiment "has the purpose of directly eliciting subjects' willingness for conditional cooperation." By comparing the responses in this experiment with those in their C-experiment, which is a standard VCM game with four-person groups (n=140), Fischbacher and Gächter argue that their results imply the vast majority of contributions are motivated by strong reciprocators. They find no evidence of interdependent utility or warm-glow altruism (no subjects stated they would contribute if other group members contributed zero). In contrast to

previous work, they claim confusion accounts for very few contributions to the public good ("at most 17.5 percent," p.3). They also argue that the decay in contributions is largely a result of interaction among free riders and strong reciprocators who revoke their cooperation once they realize they are among people who are not "norm abiders."

The small role for confusion in explaining contributions and their decay seems at odds with previous research. We believe the problem lies in the way in which strong reciprocators are identified by Fischbacher and Gächter. A strong reciprocator is any subject who states in the P-Experiment that he will contribute more if other group members contribute more. A confused, herding subject, however, would exhibit the same behavior. Thus the authors cannot discriminate between strong reciprocators and herders. Likewise, the authors claim that a positive correlation of contributions and beliefs about the contributions of others (elicited each round) is an indication of the presence of strong reciprocators, whereas it is also an indication of herders.

Thus, the six papers reviewed above offer very different conclusions about the underlying motives that generate positive contributions to the public good. Some argue that error is important and reductions in error lead to the decay in contributions over time. Others, however, argue that error is a small component of contributions and the decay stems from revocation of cooperation by conditional cooperators. The studies also come to different conclusions about the source and significance of other-regarding behavior.

We believe these differing conclusions derive from experimental designs that often change the VCM game in fundamental ways that have unknown effects on subject behavior and from empirical strategies that differ in their ability to discriminate among the hypothesized motives (none allow for all of the hypothesized motives). In the next section, we present an experimental design that addresses the aforementioned weaknesses in the current literature and attempts to resolve much of the conflicting claims about motives in VCM experiments.

Before we move on to the next section, however, we wish to address the potential role of weak reciprocity in the VCM game. Much of the recent literature ignores this motive and we believe there is good reason to do so. Andreoni (1988) developed the "Strangers" treatment in which group members in a VCM game are randomly rematched every round, as opposed to the standard "Partners" treatment in which subjects stay in the same group over rounds. Since that publication, other authors have used the Partners-Strangers design in the standard VCM game or a variant of it. A review of 15 of these experiments by Andreoni and Croson (2003) lists five

that find more cooperation among Strangers, six that find more among Partners, and four that fail to find a difference between the two treatments.

If weak reciprocity were a strong motive for behavior in the VCM, one would expect cooperation to be consistently higher in the Partners treatment. Under random rematching, weak reciprocity makes little sense. Andreoni and Croson conclude that weak reciprocity ("game theoretic effects") is unlikely to influence play in the repeated-round VCM.[5] In the analysis below, we assume that weak reciprocity is not an important determinant of VCM dynamics.

III. Methodology

We use the archetypal repeated-round, linear VCM game. Group size is four individuals who remain (anonymously) matched for a single treatment. Each subject is given an endowment of 50 laboratory tokens per round (US $0.50). The Marginal Per Capita Return (MPCR) – the marginal return from the public good to the individual relative to the value of a token kept – is constant and equal to 0.50, thus making free-riding the dominant strategy and contributing the entire endowment the socially optimal strategy. The payoff function for individual *i* is

$$\pi_i = 50 - y_i + 0.5 * (y_i + Y_i) \qquad (1)$$

where y_i is *i*'s contribution to the public good and Y_i is the contributions from the other members in *i*'s group. These attributes of the experiment are common knowledge. Instructions (see Appendix) are presented both orally and in writing. Subjects receive a payoff table that shows them the payoff from the public good ("group exchange") for every possible amount of group contributions. Every subject answers a series of practice questions that tests their understanding of payoff calculations. No subject can proceed until all the questions are answered correctly. The same author moderated all of the experiments.

After each round, subjects receive information on their contribution, the aggregate contribution of the other group members, their payoff from the group exchange, and their payoff from their private exchange (private good). On the decision screen is a "Transaction History" button, through which subjects can, at any time, observe the outcomes from previous rounds of the experiment (see Appendix for an image of the Decision Screen).

[5] The Strangers design of Fischbacher and Gächter also rules out the role of strategic considerations in motivating positive contributions in the VCM game, yet their contribution dynamics are quite typical (begin around 40% of endowment and decline to 10%).

In the "all-human" treatment, subjects play 25 rounds of the VCM game described in the first paragraph of this Section. Each subject knows that he or she will be playing 25 rounds with the same three players. To prevent individuals from discerning the identity of other group members, group assignment is random and five groups participate simultaneously in sessions of this treatment. In the "virtual-player" treatment, which is an extension of the single-shot design used by Ferraro et al. (2003), subjects play 25 rounds of the VCM game with only one important change: each human is paired with 3 virtual players (automata) and knows that he or she is grouped with virtual players. Each virtual player plays a predetermined (i.e., exogenous) contribution profile. The contribution profile is the same profile produced by a human player in a previous all-human treatment. A computer essentially scours a database of observations of human contributions in a previous all-human session and then picks at random (without replacement) a set of three human subjects from a group as the "identity" of the three virtual players. The human subjects are aware of the how the virtual players' decisions are determined. Thus each subject knows that he or she is part of a group of non-human players that behave exactly like real humans behaved in all-human groups in the same experiment.

An important feature of this design is that each human in the all-human treatment has a human "twin" in the virtual-player treatment: each twin sees exactly the same contributions by the other three members of his group in each round – the only difference is that the player in the virtual-player treatment knows he is playing with pre-programmed virtual players, not humans. Thus, for example, say subject H1 plays with H2, H3 and H4 in the all-human treatment session. Subject V1 in the virtual-player session plays with 3 virtual players, one of whom plays exactly like human subject H2, one of whom plays exactly like subject H3, and one of whom plays exactly like subject H4. This design ensures that we can treat the individual as the observational unit, rather than use the group as the independent unit of observation or make the assumption that the history of play has no effect on contributions.

To ensure that subjects in the virtual-player treatment believe the virtual player contributions are truly pre-programmed and exogenous, each subject has a sealed envelope in front of her. The subjects are told that inside the envelope are the choices for each round from the virtual players in their groups. At the end of the experiment, they can open the envelope and verify that the history of virtual group member contributions that they observed during the experiment is indeed the same as in the envelope. The subjects are informed that the reason we

provide this envelope is to prove to them that there is no deception: the virtual players behave exactly as the moderator explained they do.[6]

Each session consists of two experimental conditions, with 25 rounds of play in each. We designate the first 25-round period in a session as "*I*" (for "inexperienced") and the second 25-round period as "*E*" ("experienced"). At the beginning of each session, however, subjects are unaware that they would be playing an additional 25 rounds after the first 25 rounds. They simply begin with the instructions for the first 25 rounds. After the first 25 rounds are over, subjects are informed that there will be another 25 rounds.

Overall, with both inexperienced and experienced subject groups playing in the "all-human" (designated as "H") and virtual-player ("V") treatments, we have four experimental conditions that will be used to make inferences about the dynamics of subject behavior in the repeated-round VCM game:

1) **H*I***: Twenty-five rounds with *inexperienced*, all-human groups.
2) **V*I***: Twenty-five rounds with *inexperienced*, virtual-player groups.
3) **H*E***: Twenty-five rounds with *experienced*, all-human groups.
4) **V*E***: Twenty-five rounds with *experienced*, virtual-player groups.

The H*I* condition is the standard linear VCM game about which we wish to draw inferences about the subjects' motives. To do so, we contrast H*I* with V*I*, V*E* and H*E*. Subjects in a V*I* (V*E*) treatment observed the same history of contributions as subjects in a corresponding H*I* (H*E*) condition: each subject in H*I* (H*E*) has a "twin" in V*I* (V*E*). The only difference between H*I* (H*E*) and V*I* (V*E*) is that the humans in V*I* (V*E*) were playing with virtual players.

Since H*I* data are used in the V*I* treatment and H*E* data used in the V*E* treatment, we necessarily ran a sequence of three experiments. H*I*, by definition, had to be run first. H*I* subjects played their last 25 rounds in a modified version in which virtual-agent contributions were taken

[6] We asked two post-experiment "True or False" questions, for which our 320 subjects were paid for correctly answering: (1) The Virtual Players in your group were human beings who received money from your investment in the Group Exchange; and (2) You were able to affect how much the Virtual Players invested in the Group Exchange by changing your investment. Given the virtual-player behaviors were modeled on the behaviors of real humans, a subject might answer "True" to (1) if he misses the "your" before "investment." Only 8 subjects answered True to (1) and 3 answered True to (2). Inferences from our analysis do not pivot on the inclusion/exclusion of the subjects who answered True to these statements.

from previous H*I*, rather than H*E*, players. These last 25 rounds are thus not used in any analysis. V*I* subjects participated in V*I* and then H*E*. Only after these sessions were complete could the V*E* sessions, in which subjects first saw the V*I* contribution profile followed by virtual-agent contributions taken from H*E*, take place (otherwise any differences among experienced subjects may be due to playing with virtual players or due to subjects seeing a different history of contributions). In sum, H*I* data come from one experiment, V*I* and H*E* data from a second, and V*E* data from a third experiment.

Students from Georgia State University (Atlanta, Georgia) were recruited to participate in a computerized experiment conducted in the Georgia State University Experimental Laboratory. We have eighty subjects in each experimental condition. Subjects came from all majors and earned, on average, $33.14 for their performance in an experiment that lasted less than 1.5 hours.

IV. Analysis of Aggregate Behavior

The results of our experiments are summarized in Figure 1, which presents average contributions by round for each of the four experimental conditions. To facilitate comparisons, Table 1 presents nonparametric test statistics for selected pair-wise differences among our four conditions. For these between-subject tests we employ the Kolmogorov-Smirnov test for two independent samples (Sheskin 2000). These tests are preferred over standard *t*-tests because the distribution of contributions across subjects in a given round is very non-normal, with distinct focal points (e.g., spikes in the distribution at contributions levels of 25% and 50%) and many contributions of 0%. Statistical tests of equal contributions for selected pairs of experiment conditions are presented on a round-by-round basis as well as for the average subject-specific contributions across all rounds.

An essential maintained assumption in our analysis is that any contributions in the virtual-player treatment stem from a failure to recognize the dominant strategy of zero contributions. It is plausible, perhaps, that non-monetary considerations may have caused subjects to contribute in this treatment. Subjects may have, for example, felt compelled to contribute due to altruism towards the experiment moderator or due to a desire not to appear too greedy (Houser and Kurzban 2002). We took great care in thwarting as well as identifying these types of behavior. First, the experimenter stated that the money to pay participants came from a research grant, rather than his own pocket. Second, financial incentives were made sufficiently

large in order to establish payoff dominance: the average subject that was confused throughout the treatment forewent over $4 in earnings. Third, in a post-experiment questionnaire, we asked what the payoff-maximizing level of contributions was in the virtual treatment. Subjects were paid for correct answers (see Appendix). A comparison of these stated contributions with actual contributions provides evidence of whether subjects were indeed attempting to maximize earnings rather than attempting to please the experimenter. Using a Wilcoxon matched-pairs signed-ranks test, we fail to reject the hypothesis that stated and actual contributions are equal, using average contributions from the last five virtual-player rounds [z=0.51, p=0.61]. Overall, we feel confident that virtual-player contributions are largely due to confusion.[7]

For clarity of exposition, we organize our aggregate experimental outcome into four main results.

Result 1. Other-regarding preferences and confusion are significant motives that determine public good contributions in the standard VCM experiment (H*I*). Further, other-regarding behavior contributions *decrease* over rounds.

We focus first on comparing all-human and virtual-player contributions rates with inexperienced subjects, as this represents the cleanest distinction between contributions stemming from other-regarding motives versus those due to confusion in the standard VCM game. Contributions to the public good in H*I*, which represents the standard VCM game where inexperienced subjects play with other human subjects over repeated rounds, start at 50.1% of endowment in round 1, and steadily decline to 14.1% by round 25. This parallels the standard finding in the literature of 40 to 60% contributions in the initial period followed by a steady decline (Davis and Holt, 1993).

In comparison, V*I* contributions start at 25.6% and fall to 9.9% by round 25. On average, subjects contribute 32.5% and 19.7% of all endowments to the public good in the all-human and virtual-player treatments, respectively. Dividing V*I* contributions by H*I* contributions suggests

[7] Other data also confirm that other-regarding behavior toward the experimenter is unlikely to be an important motive. In sessions in which subjects played all 50 rounds with Virtual Players (n=80), we added two questions to the post-experiment questionnaire: "Circle the number on the rating scale that best represents your opinion about the decisions you made in the experiment. (A) I wanted to make as much money as I could for myself; (B) I wanted to make sure the professor running the experiment did not lose a lot of money." For each statement, subjects circled a number ranging from 1 (Not Important) to 7 (Very Important). The mean response to A was 6.0 and to B was 1.2 (only 11 subjects circled a number greater than "1" – 8 of them circled "2").

that 60.6% of the total contributions in the standard VCM game stem from confusion; the remaining 39.4% are attributable to other-regarding behavior. Statistical tests indicate that public good contributions are statistically *higher* in the all-human treatment at the 5% level both on average and in 24 of 25 rounds, such that other-regarding behavior is generally a statistically significant determinant.

In their closely related study, Houser and Kurzban (2002) find that, on average, 54% of the total contributions in their all-human treatment are attributable to confusion. Focusing on our first ten rounds, the length of the Houser and Kurzban experiment, our figure is 58%. These summary statistics are quite close. However, note that Houser and Kurzban find that the rate of contributions decline in the all-human treatment is statistically *slower* than the virtual-player treatment. This suggests that a larger fraction of the observed contributions is attributable to other-regarding preferences as the experiment progresses (and *less* is due to confusion). Specifically, they find that 26% of total contributions in round 1 stems from other-regarding preferences versus 73% in round 10. In contrast, our rate of decline is statistically different and *faster* for the all-human treatment (about a two-fold difference) suggesting that other-regarding behavior declines over rounds. [8] In particular, other-regarding preferences account for 49% and 48% of total contributions in rounds 1 and 10, respectively. This figure declines as our experiment progresses, with 30% of contributions due to other-regarding motives in round 25.

To put this into a different perspective, we subtract virtual-player contributions from all-human contributions and find that subjects give 24% of their endowment because of other-regarding preferences in round 1 and only give 4% of their endowment because of other-regarding preferences in round 25. The divergence in the pattern and magnitude of other-regarding preferences between our experiment and that of Houser and Kurzban may be attributable to differences in the two subject pools or due to the procedural variances in the Houser and Kurzban experimental design we highlighted previously.

Result 2. Result 1 is robust to experience.

[8] We regress mean contributions (%) on a constant and an indicator variable for the experiment round. To facilitate hypothesis tests, this is done within a time-series cross-section modeling framework (see Greene 2003, p. 320-333) whereby each treatment is a cross-sectional unit observed over a 25 period time horizon. This framework allows for treatment-specific heteroscedasticity, first-order autocorrelation, and correlation across units. The estimated relationships for the H*I* and V*I* conditions are: [H*I*] contributions = 49.05 – 1.27*round; [V*I*] contributions = 28.04 – 0.64*round. A likelihood ratio test rejects the hypothesis of equal slope coefficients for two experiment conditions [$\chi^2(1)=29.00$, $p<0.01$].

Although experienced subjects in the all-human treatment contribute less than inexperienced subjects (H*I* vs. H*E*), a finding consistent with the literature (Davis and Holt 1993), the general relationships observed between virtual-player and all-human treatments with inexperienced subjects are robust to experience (H*E* vs. V*E*). That is, there is statistical evidence that contributions stemming from other-regarding behavior are significant and are *decreasing* over rounds. In particular, other-regarding preferences account for 51%, 47%, and 25% of total contributions in rounds 1, 10, and 25, respectively. The rate of decline is approximately 1.6 times *faster* for the all-human treatment.[9]

Result 3. Contribution rates are similar across inexperienced and experienced subjects in the virtual-player treatment.

A standing hypothesis in the literature is that much of the contributions decay in the repeated-round VCM is attributable to subjects becoming aware of (i.e., "learning") the dominant strategy of zero contributions (Andreoni 1995; Palfrey and Prisbrey 1997; Houser and Kurzban 2002). The virtual-player sessions allow us to test this hypothesis because learning in this treatment is not confounded by other-regarding behavior, which may also lead to decreasing contributions in VCM games (our Results 1 and 2). Therefore, our prior expectation is that virtual-player contributions from inexperienced subjects (i.e., V*I*) would be significantly higher than subjects with prior VCM experience (i.e., V*E*). The data do not support this expectation. Average contributions are 19.7% and 11.9% of endowment with inexperienced and experienced subjects, respectively. These averages are not statistically different at the 5% level. Inexperienced subject contributions are only statistically higher than experienced subject contributions in the first three rounds and in round 22. While this pattern suggests that a few inexperienced subjects may have indeed (quickly) learned the dominant strategy, overall learning effects appear to be minimal. An alternative explanation for the decay in virtual-player contributions is that confused subjects are simply herding on the observed downward trend in virtual player contributions (which reflect behavior in past all-human sessions).

[9] Using the framework outlined in footnote 8, the estimated relationships for the H*E* and V*E* conditions are: [H*E*] contributions = 31.02 – 0.78*round; [V*E*] contributions = 18.16 – 0.48*round. A likelihood ratio test rejects the hypothesis of equal slope coefficients for these experiment conditions [$\chi^2(1)=3.88$, $p<0.05$].

Result 4. Little warm-glow or inter-dependent utility is evident in the all-human treatment.

Finally, given the standard assumption that warm-glow and interdependent utility do not decay over rounds, we can use the difference between all-human and virtual-player contributions in the last round as an upper bound on warm-glow/interdependent utility contributions. For inexperienced subjects, we have that the average subject contributes 4.2% of their endowment due to warm-glow and interdependent utility considerations. For experienced subjects, this figure is 2.3%. Putting this into another perspective, just 13.0% and 11.0% of observed contributions could be attributed to warm-glow and interdependent utility for inexperienced and experienced subjects, respectively.

V. Behavioral Model and Econometric Analysis of Individual Behavior

Behavioral Model

In this section, we develop a dynamic model of individual behavior for VCM experiments that encompasses the popular motives for public good contributions discussed in the literature. We then use econometric methods to estimate the unknown parameters of the model in order to gain insight on the relative importance of the behavioral motives under the different design conditions.

As a starting point, consider the behavior of subjects in our virtual-player treatment that do not initially deduce that their dominant strategy is to give zero contributions (and that any deviation from this behavior necessarily results in lost earnings). These "confused" individuals may look to financial signals (reinforcement learning) or to the contributions from others (herding) as indicators of optimal behavior. To incorporate these motives, we adopt a partial adjustment framework (Mason and Phillips 1997; Cason and Friedman 1999), which theorizes that the subject's decision in the current period is based on her assessment of her departure from the optimal decision in previous periods. Turning first to reinforcement learners, we depict these subjects as engaging in a hill-climbing exercise whereby their objective is to search for the profit-maximizing strategy based on financial signals from previous rounds. Let y_{it} denote individual i's contribution to the public good in round t. Further, let π_{it} denote earnings and $D_{i,t-1}$ be an indicator variable that equals 1 if the subject increases contributions from round t-2 to t-1,

equals –1 if contributions decrease between rounds t-2 and t-1, and equals 0 when contributions are unchanged. Then, a reasonable approximation is:[10]

$$y_{it} = \beta_1^{RL} y_{i,t-1} + \beta_2^{RL} y_{i,t-2} + \gamma^{RL}[D_{i,t-1}(\pi_{i,t-1} - \pi_{i,t-2})] \quad (2)$$

Inspection of this expression reveals that the reinforcement learning or "profit feedback" mechanism directs the hill climber to continue to increase (decrease) contributions if they increased (decreased) last period and earned more money or directs her to adjust contributions in the opposite direction when their last adjustment yielded lower earnings. No profit feedback is provided when contributions or profits do not change between rounds t-2 and t-1. Thus, we expect $\gamma^{RL} > 0$. We also expect γ^{RL} will be smaller in the experienced sessions, as reinforcement learning should have dissipated by then. Since current and lagged contributions should be positively correlated: $\beta_1^{RL}, \beta_2^{RL} > 0$.

Our model of herding behavior assumes that the player adjusts her contributions based on the difference between her contribution last period and the average contribution of the other group members:

$$y_{it} = \alpha^{Herd} + \beta_1^{Herd} y_{i,t-1} + \beta_2^{Herd} y_{i,t-2} + \lambda^{Herd}(y_{i,t-1} - Y_{i,t-1}/n) \quad (3)$$

For the herder, a negative (positive) deviation is a signal that she is contributing less (more) than average and should thus increase contributions. Hence, the expectation is $\lambda^{Herd} < 0$. The constant term, α^{H}, represents a baseline level of contributions the player deems optimal. It is plausible, for instance, that the player posits that she should give something, even if the virtual-players give nothing. As contributions can only be positive, expectation is $\alpha^{Herd} > 0$. Melding (2) and (3) yields our model of virtual-player treatment behavior

$$y_{it} = \alpha^{Herd} + \beta_1^{V} y_{i,t-1} + \beta_2^{V} y_{i,t-2} + \lambda^{Herd}(y_{i,t-1} - Y_{i,t-1}/n) + \gamma^{RL}[D_{i,t-1}(\pi_{i,t-1} - \pi_{i,t-2})] + \varepsilon_{it}^{V} \quad (4)$$

[10] While we include one and two-period lags of the dependent variable in our theoretical specifications, the appropriate number of lags to include (i.e., how backward-looking subjects are) is more of an empirical issue. See discussion below in the *Econometric Analysis* subsection.

where β_j^V is a weighted average of β_j^{RL} and β_j^{Herd} (for $j = 1, 2$), and ε_{it}^V is a mean-zero error term that captures the analyst's uncertainty about the specification of individual behavior.

Turning to other-regarding behavior, we consider three such motives: warm-glow, inter-dependent utility, and strong reciprocity, as defined previously. The standard assumption that warm-glow and altruism do not diminish over time (e.g., Palfrey and Prisbrey 1997) suggests the level of contributions due to either motive does not contribute to any of the observed dynamics in contributions behavior: the model of warm-glow or interdependent utility is depicted by the relationship between contributions and a constant term. Thus,

$$y_{it} = \alpha^{WG} + \alpha^{IU} \tag{5}$$

where α^{WG} and α^{IU} are specific warm-glow and interdependent utility constants, respectively.

A strong reciprocator should behave in a similar manner to a herder: she increases her contribution if the average group member is contributing more than her, and decreases contributions when she perceives she is giving too much relative to others. Thus our model of strong reciprocators is:

$$y_{it} = \alpha^{SR} + \beta_1^{SR} y_{i,t-1} + \beta_2^{SR} y_{i,t-2} + \lambda^{SR}(y_{i,t-1} - Y_{i,t-1}/n) \tag{6}$$

with expectations $\alpha^{SR} > 0$, $\beta^{SR} > 0$, and $\lambda^{SR} < 0$. An important point of emphasis is that while strong reciprocator and herder behavior may look the same, the motivation for the behavior is different. For the herding subject, the average contribution from others is a signal of how the subject should behave; for the strong reciprocator, the average contribution of the others is a signal of whether the other players are norm-abiders or they are taking advantage of the subject. Putting these other-regarding motives together yields a model of other-regarding behavior:

$$y_{it} = \alpha^{ORP} + \beta_1^{SR} y_{i,t-1} + \beta_2^{SR} y_{i,t-2} + \lambda^{SR}(y_{i,t-1} - Y_{i,t-1}/n) + \varepsilon_{it}^{ORP} \tag{7}$$

where ε_{it}^{ORP} is a mean zero disturbance term; we set $\alpha^{ORP} \equiv \alpha^{WG} + \alpha^{IU} + \alpha^{SR}$ since our experimental design does not allow us to separately identify these constant terms. Combining equations (4) and (7) we obtain our behavioral model for the all-human treatment:

$$y_{it} = \alpha^{H} + \beta_1{}^{H} y_{i,t-1} + \beta^{H}{}_2 y_{i,t-2} + \lambda^{H}(y_{i,t-1} - Y_{i,t-1}/n) + \gamma^{RL}[D_{i,t-1}(\pi_{i,t-1} - \pi_{i,t-2})] + \varepsilon_{it}{}^{H} \quad (8)$$

where β_j^{H} is a weighted average of β_j^{SR} and β_j^{V}; λ^{H} is a weighted average of λ^{V} and λ^{SR}, and $\alpha^{H} = \alpha^{Herd} + \alpha^{ORP}$. Contributions data from the all-human treatment alone do not allow one to identify the parameters α^{Herd} and α^{ORP} separately. However, estimates of these parameters are recoverable by estimating the unknown parameters of (8) and (4) with comparable all-human and virtual-player data, respectively. Since the difference between λ^{H} and λ^{C} is not equal to λ^{ORP} – the proportions of strong reciprocators and herders in the sample are not explicitly known – we cannot recover an estimate of λ^{ORP} by comparing parameter estimating from comparable all-human and virtual-player treatment data. However, larger estimates of λ with all-human treatment data indicate that strong reciprocator behavior is significant.

Econometric Analysis

In estimating the parameters of our behavioral model, it is important to account for the characteristics of our dependent variable as well as the panel structure of our data. Contributions data are discrete with a preponderance of zeros and small values. As typical in empirical work when many observations take zero values, recent efforts use Tobit models to analyze VCM data (Ashley, Ball and Eckel, 2003; Carpenter, 2004). An essential assumption of the Tobit is that zero values for the dependent variable theoretically indicate possible negative values, but these negative values are unobserved due to censoring at zero. However, this assumption is at odds with contributions data, as contributions, in principal, cannot assume negative values and zero values are *not* due to nonobservability. We instead appropriately treat our contributions data as count data and assume that the data follow a Poisson distribution. The standard Poisson maximum likelihood estimator (MLE) is still a consistent estimator of the unknown model parameters when applied to panel data (unlike the Tobit), although the standard covariance estimator is biased in this situation. To make valid inferences, we couple the Poisson MLE with White's (1982) robust covariance estimator, a.k.a. the "sandwich" estimator. We use a particular formulation of the covariance estimator for "clustered" data, which arbitrarily allows for correlation among observations from the same individual (i.e., is robust to unobserved, individual heterogeneity) and is robust to a variety of common model misspecifications (e.g., over-dispersion) (see Greene 2003). The advantage of using the Poisson MLE with a panel-

corrected covariance estimator is that one can avoid estimating a fixed or random effects model, which entails adding possibly nocuous structure to the estimator. Indeed, in employing the standard random effects MLE, the analyst assumes that the random effect is additive, uncorrelated with included regressors, and normally distributed. If, for example, the assumed distribution is incorrect, this estimator is inconsistent.

Although our behavioral model includes one and two-period lags of the dependent variable as explanatory factors, the number of lags to include (i.e., how backwards looking subjects are) is largely an empirical question. Our results are robust to alternative (i.e., higher and lower-order) specifications, and the Akaike information criterion (AIC) favors the present specification.[11] Table 2 presents estimated Poisson models corresponding to each experimental condition. Given the lagged variables included as explanatory variables in equations (4) and (8), we omit the observations from the first two rounds in each sample. We organize our econometric outcome into two main findings.

Result 5. The majority of the decline in contributions in the virtual-player treatment with inexperienced or experienced subjects arises from herding behavior. As such, there is little evidence that subjects learn the dominant strategy of zero contributions.

All parameters of the estimated models have the expected sign and are statistically significant at the 5% level, with the exception of the parameter on the profit feedback variable, which is only significant for inexperienced subjects. The lack of statistical significant on the feedback variable with experienced subjects is consistent with our expectation that most of the "hill-climbing" or "reinforcement learning" would dissipate over repeated rounds. In the interest of determining whether there is a cut-off point during the experiment where the average reinforcement learning that takes place becomes negligible, we generalized our virtual-player model for inexperienced subjects in Table 2 by allowing a structural break with respect to the feedback variable. This investigation yields an interesting result: we can reject the hypothesis that contributions due to reinforcement learning are statistically different from zero in periods 9-

[11] We estimated models (available upon request) using only a one-period lag as well as models that included up to five-period lags. Inferences drawn from these alternative specifications are similar to those presented in this paper.

25 (we can reject this hypothesis for the all-human treatment as well). Thus, it appears that the main driving force behind the decay in virtual-player contributions is herding behavior.

An investigation of the raw data reveals that the number of free riders ($0 contribution) in Round 1, Round 2, Round 24 and Round 25 of V*I* is 22, 27, 49 and 50; the corresponding numbers in V*E* are 38, 31, 46 and 58. Additional supporting evidence for Result 4 can be found in our post-experiment question about the profit-maximizing contributions level in the virtual-player treatment. Thirty percent of the subjects answered with a number greater than zero (mean = 28 tokens; median = 25 tokens).[12] Thus after 50 rounds, a substantial proportion of the subjects had not deciphered the dominant strategy. Given many subjects herded to zero contributions by Round 50, this proportion represents a lower bound on the number of confused subjects in H*I* – our focus group results (next section) suggest the proportion is much higher.

Result 6. Strong reciprocity (conditional cooperation) is a significant motive for contributions in the all-human treatment.

For both experienced and inexperienced subjects, the estimate of λ is statistically *larger* (in absolute value) in the all-human treatment than in the corresponding virtual-player treatment at the 5% significance level [inexperienced: z=1.76, p=0.04; experienced: z=2.15, p=0.02]. Thus, $\lambda^{H} > \lambda^{Herd}$, and so strong reciprocity is a significant motive for contributions in the all-human conditions.

VI. Discussion

Our analysis clearly demonstrates the substantial effects of herding and strong reciprocity on the dynamics of VCM game experiments. Thus history matters: contributions of group members in period *t*-1 influence individual contributions in period *t*. Herders look to history for a signal on how they should behave in a confusing situation. Strong reciprocators look to history to infer whether they are playing with "norm abiders" and thus whether they should continue to cooperate or begin to revoke their cooperation. Thus, analysts who model individual behavior in public goods experiments must appropriately account for the dynamics associated with repeated

[12] We did not ask this question in the first three sessions (n=60). We added the question only after being surprised by how many individuals were contributing in the last round of the virtual-player treatment.

group interactions in order to make valid inferences. Further, given the small samples used in many repeated-round experiments, simple difference of means tests between treatment and control groups can be confounded if the composition (e.g., the number of free-riders and confused individuals) of control and treatment groups differ systematically. The treatment of individuals as independent observations without controlling for group history may be one reason for conflicting experimental results in the literature.

Turning to the oft-observed decay in contributions over rounds, our analysis strongly points to interactions among free riders, strong reciprocators and herders as the main drivers of the decay. The fact that average contributions in (theoretically straight-forward) linear public goods games start between 40 and 60 percent of the endowment is consistent with the hypothesis that some participants are free riders, some are strong reciprocators, and some are initially uncertain about what to do. Ledyard (1995, p.146) conjectures that many of these uncertain subjects might simply split their endowment approximately half-half to see what happens. Our first-round data support his conjecture: in H*I*, 31 subjects chose a contribution between 20 and 30 tokens and in V*I*, 29 subjects chose a contribution between 20 and 30 tokens. Note that in H*I*, 11 subjects contributed their entire endowment, while none did in the V*I*, suggesting that most of the full-endowment contributors are not confused.

In the absence of punishment opportunities, the co-existence of free riders, strong reciprocators and herders leads to substantial decline in contributions to the public good. The initial contribution behavior, rather than the payoff outcome, starts a cascade of declining contributions through the revocation of cooperation by disappointed strong reciprocators and the herding on the downward trend by confused players. The presence of more than one herder in a group, however, may prevent universal free-riding from ever arising.

Our results imply that much of the contributions observed in VCM experiments come from confused individuals who <u>never</u> recognize the tension between the privately optimal strategy of free riding and the socially-beneficial strategy of contributing. We claim that at least 50% of observed contributions come from such subjects and at least 30% of subjects fall into this category. The latter estimate is based on the post-experiment question on the payoff-maximizing contribution, as well as the number of subjects free-riding at the end of 50 rounds of play with virtual players (V*E*). If we based our estimate on the 10^{th} round free-riding behavior of V*I* subjects, the estimated proportion of confused subjects rises to 55%. If we use 3^{rd} round

behavior, the proportion rises to 61%. Recall that we gave subjects standard VCM instructions with examples (in writing and orally), an oral summary of main features of the experiment, practice questions that had to be answered correctly before play began, and a payoff table (as well as the ability to observe the entire history of subject and group contributions).

An alternative way to categorize subjects is to examine play in rounds 10-19 in H*I* and V*I* (after hill-climbing has been abandoned). Subjects who consistently contribute at or near zero in H*I* provide an estimate of the percentage of subjects who are free riders (19%). Dividing the number of consistently positive contributors in V*I* by the number of such contributors in H*I* provides an estimate of the percentage of H*I* positive contributors who are confused (66%), which implies an estimate of the percentage of subjects who are confused (53%). The rest are strong reciprocators (28%).[13] For comparison, Fischbacher and Gächter estimate that 23% of their sample are free riders, 55% are strong reciprocators and 22% are "other."

Is it really possible that so many subjects are oblivious to the dilemma experimentalists are attempting to induce in the laboratory? To explore the question further, we paid subjects in our last session (n=20) an additional $10 to remain in the laboratory and serve as a focus group to provide feedback to the experimenters. These subjects had just completed playing 50 rounds with virtual players. Two-thirds (67.5%) of the subjects correctly answered "0" to the post-experiment question about the payoff-maximizing Group Exchange contribution. As we will see, however, many of them guessed at this answer based on the final-round behaviors of virtual players or interpreted the question as asking for the "risk-free" contribution level.

Subjects first provided written answers to six questions (see Appendix) that probed their thoughts about the experiment. Of particular interest are the answers to the question, "How did you determine how many tokens to invest in the Group Exchange in the early rounds of the experiment (first 10 rounds)." Subjects were given the following choices: (A) The choice was clear from the instructions; (B) I invested different amounts and watched how my payoff changed; (C) I observed how many tokens the Virtual Players invested and altered my decision accordingly; (D) Other (please specify). Subjects were instructed they could choose more than one response. Only 30% of subjects answered A. Fifty-five percent answered B and 65% answered C (only one subject chose D). A typical written response by a subject who contributed to the public good was, "More money could be made in the group investment versus not

[13] Using rounds 1-10, the estimates are: 10% free-riders, 66% confused and 24% strong reciprocators.

investing at all. In the previous rounds, the virtual players were on a gradual increase in investing in the group. So I wanted to get more money."

After the questionnaire was completed, the moderator asked each subject for more detail on how he or she made decisions in the "early" rounds (first 10 rounds) of the experiment. The order in which subjects were questioned was determined by the monitor's observations of the data from rounds 6 – 20. The order was based loosely on how confused the subjects appeared to the moderator, which was determined by their contribution patterns. Subjects who persisted in making positive contributions or frequently changed their contribution levels were considered more confused and subjects who generally contributed zero were labeled least confused. We ordered subjects in this way to mitigate the risk that confused subject responses would be affected by the responses of subjects who understood the incentives.

Only 25% of the subjects said that the payoff-maximizing strategy was clear from the instructions (some contributed a few tokens now and then just to confirm their understanding of the game). Ten percent of subjects reported having no idea about what was going on and simply chose contribution levels at random. Another 10% attempted, without success, to vary their contributions and infer a pattern. Twenty percent reported depending solely on the behavior of the virtual players to determine their own contribution.

Thirty-five percent of the subjects reported a mix of beginning with a split of their endowment, followed by watching what the virtual players were doing and by attempting to infer if there was any pattern to earnings, followed quickly by abandoning any attempt to infer a pattern and instead herding along with the virtual players. Only one of these subjects reported finally "getting it" and changing his behavior for the second set of 25 rounds. Thus, as implied by our econometric results, some subjects attempted to infer the best response strategy from play of the game, but found it too difficult, gave up and simply imitated what they saw other players doing. In retrospect, this result is not surprising. If a subject was unable to see from the instructions that every token invested in the Group Exchange yielded him only one-half token, the same subject is unlikely to make the inference from observing changes in earnings when his contributions and those of his group members were changing simultaneously.

Recall that two-thirds of the subjects answered that contributing zero tokens to the Group Exchange would maximize their payoffs. When asked why they wrote down zero, but did not invest zero, two general responses were heard: (1) one had to come up with an answer and given

the virtual players were contributing at zero or near zero in the final rounds, an answer of "0" seemed like the best answer; and (2) the question was asking about the "risk-free" investment decision. This latter response was common, orally and in writing, among self-reported herders. When probed, many subjects spoke of a perceived "risk" associated with investing in the Group Exchange. As the following two written answers imply, many subjects understood that higher group payoffs were engendered when all members contributed, but they mistakenly thought that this outcome maximized their own earnings.

> "If I wanted to play it safe, I would invest nothing at all. But in order to maximize my earnings, every member (including virtual players) would need to invest."
>
> "Put 50 in Individual and 0 in Group. This would mean your money is guaranteed. The other option is risky."

The oral discussion suggests there are two types of herders: (1) the majority of herders who are confused and just follow average contributions of group; and (2) others who have a more sophisticated, but incomplete, understanding of the game in which they are playing. They incorrectly believe that it is privately optimal to contribute more when others are contributing more, and contribute less when others are contributing less. They have a sense of being "suckers" if they contribute and their group members do not, but they do not understand, even after 50 rounds of play with virtual players, that they would be better off by free-riding on the other group members' contributions. These players seem to view the game as an assurance game, rather than a linear public goods game.

The oral response of one subject captures the sentiment of this sub-group: "The way to maximize your earnings was to invest when the Virtual Players invested and don't invest when they didn't invest. I would have made a lot more money if I had been with other Virtual Players. The ones I had in the second 25 periods were jerks. In the first 25, the virtuals invested a lot more in the Group Exchange than the ones I had in the last 25 rounds. I hardly invested anything in the Group Exchange with the last group." The moderator asked her, "So if your Virtual Players had invested 50 tokens every period, you would also have invested 50 tokens?" She said, "Yes, that would have ensured I made the most money." She then pointed to her payoff table and stated that more money was made when more tokens were invested. Note that this woman (1) understood she was playing with robots whose behavior she could not change, (2)

correctly answered the post-experiment question on the payoff-maximizing contribution and (3) had just listened to another subject articulate the dominant strategy in precise terms.

Thus the empirical results and the *ex post* subject narratives (transcripts available upon request) all suggest that a substantial proportion of subjects begin and end the experiment without recognizing the tension between the privately optimal strategy of free-riding and the socially-beneficial strategy of contributing to the public good.

VII. Conclusion

Understanding behavior in experimental implementations of the Voluntary Contributions Mechanism game is environmental economists with institutional and policy-oriented interests. We argue that the dynamics typically observed in VCM experiments are a result of interactions among free riders, strong reciprocators and confused herders. Contrary to widely held beliefs, we find little evidence that subjects who initially are unable deduce the dominant strategy of zero public good contributions ever "learn" this strategy through repeated play. These results have important implications for theoretical and empirical research and for the external validity of VCM experiments.

What are the implications of strong reciprocity for theoretical and empirical research? The existence of strong reciprocators implies more complicated interactions among agents than have previously been assumed in public goods games. Beliefs and history are key determinants to the behavior of strong reciprocators. Ignoring such factors will lead to erroneous theoretical predictions and misplaced inferences based on empirical data. A better understanding of the behavior of strong reciprocators is an important area for future research.

The existence of strong reciprocators also has environmental policy implications. The effects of a policy may be different if many people are strong reciprocators rather than impure or pure altruists. For example, common knowledge of extensive free-riding in the provision of an environmental good (e.g., ignoring water restrictions, tampering with pollution control devices) can be disastrous for a society populated with strong reciprocators (and herders). Likewise, common knowledge of extensive environmental compliance may result in the perseverance of high levels of compliance over time. Such knowledge, however, would not have any effect on behavior if those who contribute to supplying environmental goods were motivated by pure or impure altruism.

This dependence on group dynamics may be one reason that conservation education rarely seems to have long-term effects on conservation outcomes. Although such education can provide a social norm around which individuals can cooperate, free-riding by even a minority can eventually lead to a collapse of cooperation by those who believe the social norm is worthwhile. Moreover, the common practice of environmental awareness campaigns to emphasize how many citizens are free-riders only encourages more free-riding from strong reciprocators (and herders). A more appropriate strategy would be to emphasize the average contributions from contributors and not announce the frequency of free riders.

What are the implications of herding behavior for theoretical and empirical research? Clearly, such behavior implies a specific structure on individual decision-making errors and this structure can be incorporated into theoretical and empirical models. More difficult to discern, however, is what such error implies about the ability of experimental economists to make predictions about behavior in non-experimental settings. For instance, in naturally-occurring public goods situations, people may not know how much to contribute to a public good and will look to neighbors or others for guidance. However, we believe that individuals in such situations *do* recognize the tension between the privately optimal strategy of free riding and the socially-beneficial strategy of contributing some positive amount. They look to others for a signal of a value of the public good. Herding in public good experiments may thus be fundamentally different from herding in naturally-occurring public goods situations. Herding in public good experiments may be a result of the opaque, "neutral" language of experimental instructions, rather than a relevant empirical phenomenon for economists.

The presence of herders may thus pose a problem for those who wish to use public good experiments to make predictions about human behavior in environmental contexts. Laury and Taylor (2004) use behavior in a one-shot VCM game experiment to predict behavior in a situation in which individuals can contribute to a naturally-occurring environmental public good (urban tree planting). Using the empirical approach of Goeree, Holt and Laury to estimate an altruism (interdependent utility) parameter for each subject, the authors find little or no relationship between subjects' altruism parameters and subjects' behaviors in the naturally-occurring public good situation. Such a result could arise if herding behavior mimics the behavior of altruists in their VCM experiment: for example, the variations in MPCR used to identify altruistic motivations could also serve as external signals of appropriate behavior for

herders. Their confusion, however, would not be incorporated into the Logit equilibrium model's noise parameter, but rather the altruism parameter.[14]

In addition to shedding much needed light on what is going on in VCM experiments, our analysis contributes to the growing body of evidence that suggests there are different "types" of economic agents and the interaction of these agents gives rise to the observed dynamics of behavior and economic outcomes in our world. More collaboration by theorists and experimentalists to understand agent types and model their interactions would be extremely fruitful in helping economists explain how humans, rather than hyper-rational automata, behave in the games in which they play.

[14] We have preliminary data in which we replicated the Goeree, Holt and Laury experiment with all-human groups and human-virtual player groups. Our data are consistent with the idea that the altruism coefficients estimates result from error, not altruism.

Table 1. Nonparametric, Kolmogorov-Smirnov Difference Tests[1, 2]

Round	H*I* v. V*I*	H*E* v. V*E*	H*I* v. H*E*	V*I* v. V*E*
1	0.3125**	0.3000**	0.2625**	0.2500**
2	0.3250**	0.2000*	0.3750**	0.2125*
3	0.2750**	0.1250	0.3250**	0.2000*
4	0.2750**	0.1000	0.3750**	0.1625
5	0.2750**	0.1250	0.3250**	0.1625
6	0.3375**	0.2000*	0.2375*	0.1250
7	0.3500**	0.2500**	0.3000**	0.1625
8	0.3250**	0.0625	0.3375**	0.1125
9	0.3875**	0.1125	0.3875**	0.1125
10	0.3625**	0.2000*	0.3250**	0.1500
11	0.3500**	0.2250*	0.2125*	0.1250
12	0.3875**	0.2875**	0.2125*	0.1375
13	0.2500**	0.2000*	0.2125*	0.1375
14	0.3250**	0.2000*	0.1875	0.1125
15	0.3250**	0.1750	0.1875	0.0625
16	0.3125**	0.2000*	0.2125*	0.1750
17	0.3250**	0.2750**	0.2250*	0.1500
18	0.2750**	0.2125*	0.1375	0.1125
19	0.3250**	0.2000*	0.2125*	0.0625
20	0.1875	0.2375*	0.1250	0.1625
21	0.3125**	0.1375	0.2875**	0.1125
22	0.2750**	0.1250	0.2875**	0.2000*
23	0.3500**	0.1000	0.3375**	0.1375
24	0.2000*	0.1000	0.1500	0.0500
25	0.2000*	0.0500	0.2625**	0.1250
Average	0.4125**	0.2500**	0.3375**	0.1750

* and ** correspond respectively to 5%, and 1% significance levels (one-sided tests).

[1] "H" and "V" refer to the All-Human and Virtual-Player treatments, respectively. "*I* " and "*E* " refer to inexperienced and experienced subjects, respectively.

[2] Critical values for one-tailed Kolmogorov-Smirnov *D*-statistics are: 5% = 0.1929; 1% = 0.2403.

Table 2. Dynamic Poisson Models of Individual Behavior

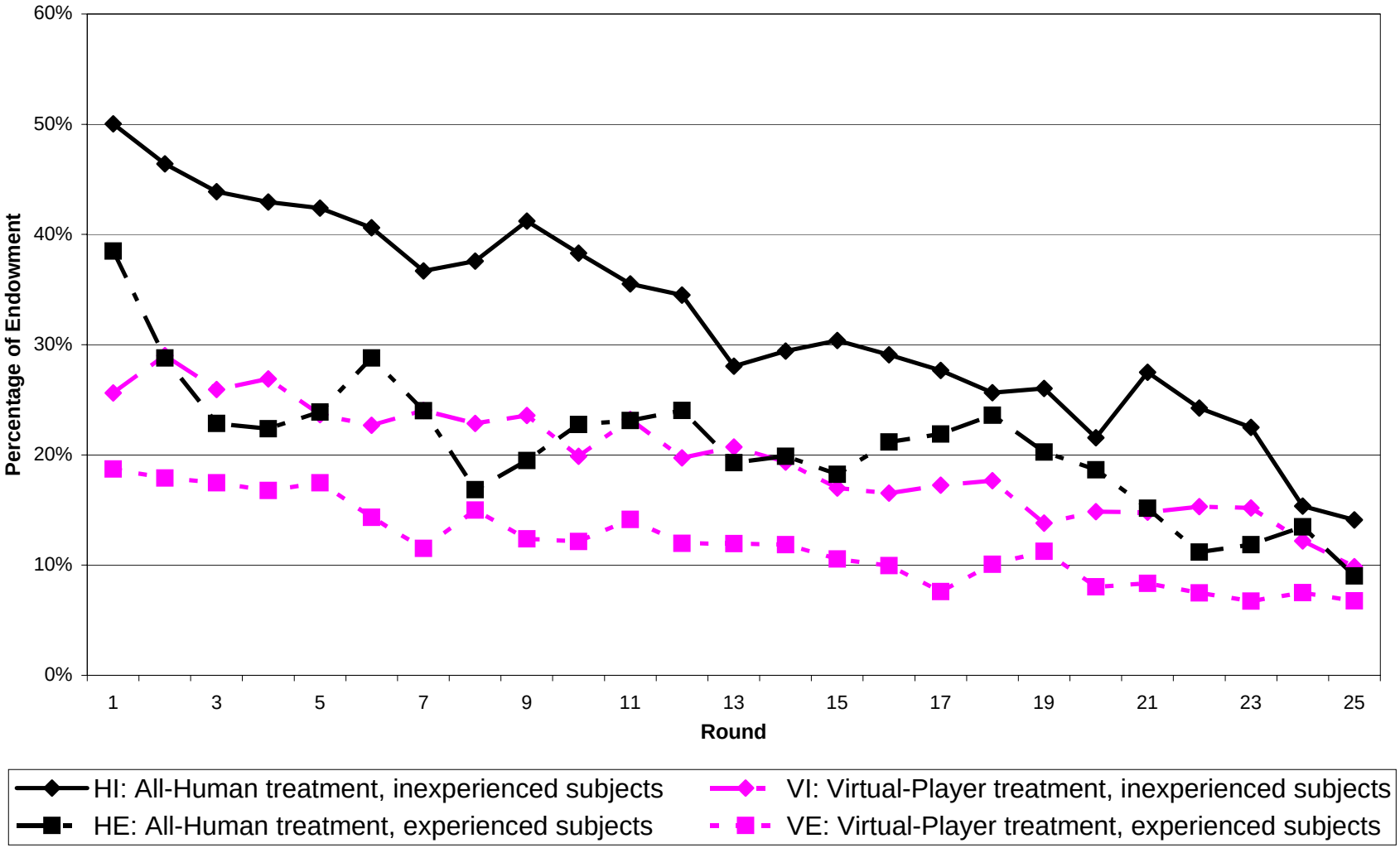
Dependent variable is y_{it} (i's contribution to the public good in round t)

Variable	Parameter	All-Human, inexperienced	Virtual-Player, inexperienced	All-Human, experienced	Virtual-Player, experienced
Intercept	α	1.8516 (0.0738)**	1.2482 (0.1353)**	1.3803 (0.1001)**	0.9379 (0.1235)**
$y_{i,t-1}$ [subject contributions in round t-1]	β_1	0.0337 (0.0028)**	0.0376 (0.0047)**	0.0497 (0.0037)**	0.0462 (0.0059)**
$y_{i,t-2}$ [subject contributions in round t-2]	β_2	0.0141 (0.0016)**	0.0298 (0.0034)**	0.0150 (0.0024)**	0.0353 (0.0038)**
$y_{i,t-1} - (Y_{i,t-1}/n)$ [deviation from average contributions of other group members in round t-1]	λ	-0.0153 (0.0023)**	-0.0072 (0.0040)*	-0.0256 (0.0035)**	-0.0143 (0.0039)**
$D_{i,t-1}(\pi_{i,t-1} - \pi_{i,t-2})$ [profit "feedback" mechanism]	γ	0.0044 (0.0018)**	0.0062 (0.0033)*	-0.0003 (0.0037)	0.0039 (0.0045)
Log-Likelihood		-13,751.38	-12,538.29	-13,530.07	-10,134.32
N		1840	1840	1840	1840

Note: standard errors are in parentheses.

* and ** indicate that parameters are statistically different from zero at the 5% and 1% level, respectively. Consistent with our theoretical hypotheses, these are one-sided tests.

Figure 1. Mean Contributions per Round by Experiment Condition



References

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

Andreoni, J. 1990. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal* 100(401): 464-477.

Andreoni, J. 1995. Cooperation in Public-goods Experiments: Kindness or Confusion? *American Economic Review* **85**(4): 891-904.

Andreoni, J. and R. Croson. 2003. Partners versus Strangers: Random Rematching in Public Goods Experiments. In V. Smith and C. Plott, eds., *Handbook of Experimental Economics Results*.

Ashley, R. S. Ball, and C. Eckel. 2003. Analysis of Public Goods Experiments Using Dynamic Panel Regression Models. Working Paper, Department of Economics, Virginia Tech.

Bowles, S., E. Fehr and H. Gintis. 2003. Strong Reciprocity May Evolve With or Without Group Selection. *Theoretical Primatology* (December): 1- 8.

Carpenter, J. 2004. When in Rome: Conformity and the Provision of Public Goods. *Journal of Socio-Economics* **33**(4): 395-408.

Cason, T.N. and D. Friedman. 1999. Learning in Laboratory Markets with Random Supply and Demand. *Experimental Economics* **2**(1): 77-98.

Davis, D.D. and C.A. Holt. 1993. *Experimental Economics*. Princeton, N.J.: Princeton University Press.

Ferraro, P.J., D. Rondeau, and G.L. Poe. 2003. Detecting Other-regarding Behavior with Virtual Players. *Journal of Economic Behavior and Organization* **51**: 99-109.

Fischbacher, U. and S. Gächter. 2004. Heterogeneous Motivations and the Dynamics of Free Riding in Public Goods. Working Paper, Institute for Empirical Research in Economics, University of Zurich.

Goeree, J., C. Holt, and S. Laury. 2002. Private Costs and Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior. *Journal of Public Economics* **83**: 255-276.

Greene, W.H. 2003. *Econometric Analysis*, fifth edition. Upper Saddle River, N.J.: Prentice Hall.

Houser, D. and R. Kurzban. 2002. Revisiting Kindness and Confusion in Public Goods Experiments. *American Economic Review* **92**(4): 1062-1069.

Isaac, R.M., J. Walker, and S. Thomas. 1984. Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations. *Public Choice* **43**: 113-149.

Laury, Susan and Laura Taylor. 2004. Altruism Spillovers: Does Laboratory Behavior Predict Altruism in the Field? Working Paper. Georgia State University, Atlanta, GA.

Ledyard, J. 1995. Public Goods: A Survey of Experimental Research. In J.H. Kagel and A.E. Roth eds, *Handbook of Experimental Economics*. Princeton, Princeton University Press, pp. 111-194.

Mason, C.F. and O.R. Phillips. 1997. Mitigating the Tragedy of the Commons through Cooperation: An Experimental Evaluation. *Journal of Environmental Economics and Management* **34**: 148-172.

Palfrey, T.P. and J.E. Prisbrey. 1997. Anomalous Behavior in Public Goods Experiments: How Much and Why? *American Economic Review* **87**: 829-846.

Sheskin, D.J. 2000. *Handbook of Parametric and Nonparametric Statistical Procedures*, second edition. Boca Raton, Florida: Chapman and Hall.

White, H. 1982. Maximum Likelihood Estimation of Misspecified Models. *Econometrica* **50**(1): 1-25.