Stylized Facts and Identification in Public Goods Experiments:

The Confusion Confound

Paul J. Ferraro
Department of Economics, Andrew Young School of Policy Studies,
Georgia State University, P.O. Box 3992, Atlanta, GA 30302-3992
Phone: 404-651-1372; Fax: 404-651-0425; E-mail: pferraro@gsu.edu

Christian A. Vossler
Department of Economics, 534 Stokely Management Center,
The University of Tennessee, Knoxville, TN 37996-0550
Phone: 865-974-1699; Fax: 865-974-4601; E-mail: cvossler@utk.edu


* Research supported by the Georgia State University Office of Sponsored Programs.
Thanks to Krawee Ackaramongkolrotn and Steve Cotten for software development, James Cox, Susan Laury and the GSU Research Initiation Grant for funding, Kwaw Andam for data entry, and Douglas Davis, Catherine Eckel, Joseph Henrich, Susan Laury, Ragan Petrie and participants at the 74th Annual Meeting of Southern Economic Association, and 2004 North American Regional Meetings and 2006 International Meetings of the Economic Science Association for useful comments.
Stylized Facts and Identification in Public Goods Experiments: The Confusion Confound

Abstract

Using a novel experimental design and microeconometric modeling, we demonstrate that the stylized facts of twenty-five years of public good experiments may be artifacts of the instructions and designs commonly used by experimentalists. Although previous studies have identified confusion as a contributions motive, we establish that the behavior of confused subjects amounts to much more than statistical noise, and does not dissipate with repetition as previously claimed. Confused subjects use experimental parameters and the behavior of other players as behavioral cues, which confounds traditional strategies to identify other-regarding preferences though exogenous parameter changes and the modeling of reactions to other subjects’ decisions. Although we find evidence of pure altruism and conditional cooperation, we are unable to draw inferences about a substantial proportion of subjects who either are unable to infer any clear incentives in the game or erroneously believe they are playing an Assurance Game with multiple equilibria and no dominant strategy.

Keywords: Public goods; experimental design; conditional cooperation; herding; dynamic modeling

JEL classification: H41, C90, C72, C20
I. INTRODUCTION

The Voluntary Contributions Mechanism (VCM) game is central to experimental research on the private provision of public goods. Public economists use the VCM to test their 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.

The typical linear VCM experiment places subjects in a social dilemma. 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 group member, 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 individual has a dominant strategy to free-ride and all players contributing zero tokens to the public account represents a unique Nash equilibrium. The Pareto-dominant, welfare-maximizing outcome, however, is realized when everyone contributes their entire endowment to the public account.

Over 25 years of VCM experiments have resulted in the following stylized facts (Davis and Holt, 1993; Ledyard, 1995; Holt and Laury, forthcoming):

(1) Average contributions to the public good are a significant portion of endowments.

   (a) In single-shot settings, average contributions are between 40-60% of endowments.

   (b) In repeated-round settings, average contributions start in the same range, but decline over rounds and remain significantly different from zero in the last round (even in experiments that last 50 or more rounds).

(2) Increases in the Marginal per Capita Return (MPCR) increase contributions.
(3) Increasing group size, at least for low MPCRs and small group sizes, increases contributions.

Standard game theory that assumes purely self-interested players does not explain observed contributions in single-shot or repeated-round settings, and thus alternative theories and supporting empirical evidence have been put forth. Yet, there is no consensus on what is driving the results of public goods games (Ledyard, 1995). At least some results are consistent with other-regarding behaviors: warm-glow altruism (e.g., Palfrey and Prisbey, 1997), pure altruism (e.g. Goeree, Holt and Laury, 2002), conditional cooperation (e.g., Fehr, Fischbacher and Gächter, 2001; Fischbacher and Gächter, 2004), or a combination of these (e.g., Cox and Sadiraj, 2005).\(^1\)

The standard contributions decline in repeated-round settings has been attributed to reductions in confusion (Andreoni, 1995; Palfrey and Prisbrey, 1997; Houser and Kurzban, 2002; Carpenter, 2004), or to the revocation of conditional cooperation (Fishbacher, Gächter and Fehr, 2001, and others who use the strategy method).

This study focuses on the role of confusion behavior. We use the term “confusion” to characterize behavior that stems from subjects’ inability to discern the nature of the game in which they are playing. Such players, for example, are unable to discern the dominant strategy of zero contributions. Past studies find that confusion is a source of a significant fraction of contributions, as much as 50% (Andreoni, 1995; Palfrey and Prisbrey, 1997; Houser and Kurzban, 2002). This result is troubling, as from the perspective of an economist at least, the VCM is a rather simple game. Moreover, there is likely no direct analog to this type of behavior.

---

\(^1\) In this article, “other-regarding behavior” will be used as an umbrella term to characterize three motives for contributions in the VCM game: (1) “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 altruism” (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) “conditional cooperation” (sometimes referred to as “strong reciprocity”), which describes a situation where participants try to match or condition their contributions to the contributions of their group members.
in a naturally occurring setting: we hope that when presented with a contribution decision, an individual only contributes if this increases her utility. The proportion of contributions stemming from confusion in this study is similar to previous studies. However, we investigate the confusion phenomenon more intricately to reveal that confusion confounds studies that strive to identify the relative importance of various motives.

We present data generated from three distinct experimental designs, chosen to be representative of the literature, in conjunction with a “virtual-player method” to separate contributions stemming from other-regarding preferences from those due to confusion. These experiments allow us to demonstrate robustly that the stylized facts of twenty-five years of public good experiments may be artifacts of the instructions and designs commonly used by experimentalists. Our results support recent claims that methodological shortcomings decrease the external validity of laboratory experiments (Loewenstein, 1999; Levitt and List, 2006). To be clear, we are not saying that stylized facts do not exist, but rather that we cannot make inferences for a large proportion of subjects in these studies. In the absence of confusion the stylized facts may or may not be considerably different.

Experiment 1 uses the design of Goeree, Holt and Laury (2002) to show that confusion behavior confounds strategies to identify other-regarding preferences through the use of exogenous parameter changes. In this experiment, confused subjects behave like pure altruists. Experiments 2 and 3 show that confusion behavior, in repeated-round and single-shot settings, confounds strategies to identify other-regarding preferences though the modeling of reactions to other subjects’ decisions. Results from Experiment 2, a standard repeated-round VCM game (e.g. Isaac, Walker and Thomas, 1984), show that confusion does not dissipate with repetition and that some confused individuals behave like conditional cooperators. Results from a post-experiment
focus group reveal that part of the confusion stems from some subjects introducing their own context to the game, treating it instead as an Assurance Game, which is a type of coordination game under which players do not construe zero contributions as free-riding, but as the risk-dominant strategy. Experiment 3 uses the increasingly used ‘strategy method’ design of Fishbacher, Gächter and Fehr (2001) to demonstrate that many confused individuals in these experiments are erroneously classified as conditional cooperators.

II. THE VIRTUAL-PLAYER METHOD

The virtual-player method discriminates between confusion and other-regarding behavior in single-round public goods experiments (Ferraro, Rondeau and Poe, 2003), and discriminates between confusion and other-regarding behavior or self-interested strategic play in repeated-round experiments. The method relies on three important features: (1) the introduction of nonhuman, virtual players (i.e., automata) that are programmed to play decisions identical to those made by human players in an otherwise comparable treatment; (2) a split-sample design where each participant knowingly plays with either humans (the “all-human treatment”) or with virtual players (the “virtual-player treatment”); and (3) a procedure that ensures that human participants understand how the nonhuman, virtual players behave. In the repeated-game context, the method also ensures that subjects in the all-human and virtual-player treatments see the same history of play by their group members.

Here is an excerpt from the virtual-player instructions for Experiment 1 (Section III):

“In this experiment you will be asked to make a series of choices about how to allocate a set of tokens. You will be in groups for this experiment. However, you will not be grouped with others in the room. Your group will consist of yourself and “Virtual Players.” These Virtual Players are not human and their decisions have already been determined. Your decisions will thus have absolutely no effect on how the Virtual Players behave. To assure you that the decisions of the nonhuman Virtual Players have indeed
been determined already and will not change during the experiment, we have envelopes in which the investment decisions of the nonhuman Virtual Players in your group are printed on a piece of paper. We have placed these envelopes on your desk. AFTER the experiment is over, you may open your envelope and confirm that it contains the decisions made by the nonhuman Virtual Players in your group. PLEASE DO NOT OPEN THE ENVELOPE UNTIL THE EXPERIMENT IS COMPLETED.”

Our other experiments use similar language (see appendices for instructions). Using virtual-player language allows us to use traditional experimental instructions and thus test hypotheses without changing the nature of the game. The instructions emphasize the nonhuman nature of the virtual players and the exact way in which virtual player decisions are programmed. To avoid misunderstandings in payoff calculation examples and practice questions, we are careful to place words such as “earn” between quotes when describing payoffs for the virtual players and then immediately insert further explanation such as, “*Of course, because the Virtual Player is not real, it does not actually receive any earnings.*” To assure subjects that there is no deception, we give each subject a sealed envelope containing virtual player decisions, which they are free to open at the end of the experiment.

The random assignment of participants to an all-human group or a virtual-player group allows the researcher to net out confusion contributions by subtracting contributions from (human) participants in the virtual-player treatment from contributions in the all-human treatment. In single-round experiments where the decisions of other players are not known *ex ante*, the contributions from virtual players should have no effect on human contributions nor should they confound any comparison between all-human and virtual-player treatments. Thus, randomly selecting the profile of any previous human participant, with replacement, as the contribution profile for a virtual player suffices to ensure comparability.

---

2 Johnson et al. (2002) use nonhuman players in a sequential bargaining game, but these players are programmed to play a subgame perfect equilibrium strategy rather than as humans have played in previous experiments. Houser and Kurzban (2002) also used nonhuman players, but as we describe in Section IV, their design differs in several important ways.
In repeated-round public goods games where group contributions levels are announced after each period, one must exercise additional control because the history of play may affect contributions. The additional control comes by ensuring 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 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 contributes exactly like human subject H2, one of whom contributes exactly like subject H3, and one of whom contributes 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.

Participant Comprehension and Efficacy of the Virtual Player Conditions

When a subject knowingly plays a linear VCM game with automata whose contributions are unaffected by the subject, we maintain that the only significant motive for contributing is the inability to infer the dominant strategy. Below, we establish that in virtual-player treatments: (1) subjects clearly understand the nature of virtual players; (2) there are no important motives other than confusion for contributions; and (3) subjects behave as if profit-maximization is their primary motive.

First, we asked “True or False” questions about the nature of the virtual group members (e.g., in Experiment 2, subjects were asked: “The members of your group were human beings who received money from your investment in the Group Exchange) and the exogeneity of the decisions made by the group members (e.g., “You were able to affect how much the Virtual
Players invested in the Group Exchange by changing your investment.”). In Experiments 1 and 2, we asked these questions after subjects had made their decisions. Because of concerns that post-experiment questionnaires do not necessarily capture pre-decision understanding, we asked this question before subjects made their decisions in Experiment 3 (no subject was allowed to continue if they answered the question incorrectly).

In Experiment 1, three virtual-player treatment participants (6%) answered that their group members were human and three (6%) answered that they could affect the decisions of their virtual player group members. For Experiment 2, these same figures are just 3% and 1%. Thus, evidence from the questionnaire strongly suggests that participants understood the role of virtual players (and our results do not pivot on the inclusion/exclusion of the few participants who answered the questions incorrectly).

Considerations other than own payoff-maximization could cause subjects to contribute. Subjects may, for example, feel compelled to contribute due to altruism towards the experiment moderator or due to a desire not to appear too greedy (Harrison and Johnson, forthcoming). We took great care to thwart these motivations. First, the experimenter stated that the money to pay participants came from a research grant, rather than his own pocket. Second, the participants are repeatedly told that virtual players do not receive earnings from public good contributions. Third, we use financial incentives to establish payoff dominance: the difference in earnings between contributing the entire endowment and contributing nothing is about $3.25 in Experiment 1 and $6.25 in Experiment 2 (for 25 rounds). In Experiment 2, for example, the average confused subject forewent over $4 in earnings over 25 rounds.

---

3 They may have believed the question was asking about the source of the virtual player contributions, which was human, rather than the nature of the virtual players, which was nonhuman.
4 In Experiment 3 (virtual-player only), one subject answered the “human” T/F question incorrectly and when prompted to reread the question, changed his answer. Another subject answered the “pre-determined” T/F question incorrectly and when prompted to reread the question, changed her answer.
In sessions where subjects \( (n=80) \) played 50 rounds of the VCM game with Virtual Players, we posed two questions after the experiment: “Circle the number on the rating scale that best represents your opinion about the decisions you made in the experiment. (C) I wanted to make as much money as I could for myself; (D) 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 C was 6.0 and to D was 1.2 (only 11 subjects circled a number greater than “1”; 8 of them circled “2”).

For the last session in which subjects played 50 rounds with virtual players \( (n=20) \), we conducted an in-depth focus group in which we probed, in writing and then orally, subject understanding of the incentives and their motivations for contributing or not contributing. The results are detailed in Section IV and the appendix, but the bottom-line is that the focus group revealed that subjects clearly understood the nonhuman nature of the virtual players and their inability to change the virtual players’ decisions.

Furthermore, after Experiments 1 and 2, participants were asked to identify the payoff-maximizing level of contributions in the virtual treatment. Subjects were paid for correct answers (see appendix). Since Experiment 1 included decision tasks with an MPCR of 1, we asked for the profit-maximizing contribution associated with a particular task in this experiment \( (#15) \). A comparison of these stated contributions with actual contributions in the experiment suggest whether or not subjects were indeed attempting to maximize earnings rather than attempting to, say, please the experimenter. Using a Wilcoxon matched-pairs signed-ranks test, we fail to reject the hypothesis that stated and actual contributions are equal in Experiment 1 \( [z = 0.65, p = 0.51] \), or, using average contributions from the last five virtual-player rounds, in Experiment 2 \( [z = 0.51, p = 0.61] \). In sum, we are confident that virtual-player contributions are
largely due to confusion over the dominant strategy of zero contributions, and not due to an incomplete understanding of the role of virtual players or non-monetary motives.

III. EXPERIMENT 1: ALTRUISM AND THE EFFECTS OF MPCR AND GROUP SIZE

Experimental Design

In order to detect contributions stemming from pure and warm-glow altruism, Goeree, Holt and Laury (2002) (hereafter GHL) create a variant of the one-shot linear VCM game and empirically model contributions using a quantal response equilibrium model of noisy decision-making. Each subject decides how to allocate 25 tokens between a private and a public account in each of ten “one-shot” decision tasks, without feedback, where the internal ($m_I$) and external rates ($m_e$) of return, and group size ($n$), vary across tasks. The internal rate of return is the marginal return to oneself from a token contributed to the public account. The external rate of return is the marginal return to other players from one’s contribution to the public account. Group size is either two or four players. The internal rate of return is always lower than the value of a token in the private account, and thus subjects have a dominant strategy to contribute nothing to the public account. Subjects know they will be paid for only one of their decisions, chosen at random ex post.

In the typical one-shot VCM game, all players receive the same return from the public good (i.e., $m_I = m_e$). Varying the returns and group size serves to identify pure altruism in their empirical model: participants exhibiting pure altruism should increase their contributions when the external return or the group size increases.5 If considerable contributions are observed, but they show little correlation with external return and group size, the conjecture is that contributions are largely attributable to warm-glow altruism. We further hypothesize that

5 Palfrey and Prisbey (1997) also use exogenous changes in parameters to identify altruism.
individuals who are confused about the incentives in the game will believe the variation in the returns and group size is a signal about the payoff-maximizing allocation of tokens.

We thus create an augmented GHL design, in which we add five additional decision tasks that allow us to generate greater variability in the MPCR and group size. We also increase the financial returns from private and public account allocations (GHL earnings are so low that an unrelated experiment was needed to augment earnings). Endowments are 50 tokens per task, private account returns are 6 cents/token, internal returns from the public good are 1, 2, 4 or 6 cents, and external returns from the public good are 1, 2, 4, 6 and 12 cents.

Experiment instructions are presented orally and in writing (see appendix A). The all-human treatment instructions are from GHL, with minor changes. The virtual-player instructions are similar with the exception of emphasizing that participants are matched with pre-determined contributions. As in GHL, participants make decisions via paper and pencil and must answer a series of practice questions before making their decisions. A post-experiment questionnaire is given to collect basic demographic information as well as to assess understanding of the experimental design and decision tasks. The same author moderated all of the sessions.

GHL assign subjects to two- and four-member groups by selecting marked ping-pong balls after all decisions are made. We pre-assign participants to two- and six-member groups based on their subject ID number. This pre-assignment is important for virtual-player sessions because we give each subject a sealed envelope containing information on the aggregate contributions of the virtual players before the subjects make their decisions.

Ninety-six subjects were recruited from a pool of undergraduate student volunteers at the University of Tennessee (Knoxville). The sample was split between the all-human and virtual-player treatments. Sessions consisted of twelve people, and participants were visually isolated
through the use of dividers. Subjects were not aware of the identity of the other members of their groups. All sessions took place in the University of Tennessee Experimental Economics Laboratory. Earnings averaged $14.98 and the experiment lasted no more than one hour.

Results

Figure 1 presents mean contributions (as a percentage of endowment) in the two treatments for each of the fifteen decision tasks. Looking at the all-human treatment, the pattern of contributions in relation to design factors is quite similar to the GHL study, with contributions increasing with respect to external return, group size, and MPCR. Note, however, that the same pattern, albeit at a lower level, is observed in the virtual-player treatment. Subjects in the virtual-player treatment alter their contributions based on the same stimuli as subjects in the all-human treatment: the two response patterns are parallel. Contributions are nearly equal only for the two decision tasks in which the MPCR is equal to one: most subjects in the virtual-player sessions who contributed zero for all other decision tasks contributed full endowments when it cost them nothing to do so (a clear indication that they understood the payoff incentives).

Result 1. About half of contributions to the public good stem from confusion.

Excluding treatments where the MPRR equals one, on average public good contributions are 24.1% of endowment in the all-human treatment. This figure is 12.0% in the virtual-player treatment. The ratio of virtual-player contributions to all-human contributions provides an estimate of the proportion of all-human treatment contributions that stem from confusion: 12.0% / 24.1% or 49.9%. An estimate of approximately half of contributions to the public good resulting from confusion is consistent with previous results (Andreoni, 1995; Houser and

---

6 The order of decision tasks in the experiment differs from the order of presentation in the figure.
Kurzban, 2002; Ferraro, Rondeau and Poe 2003), and consistent with results from Experiment 2, but at odds with recent research on conditional cooperation.

**Result 2.** Contributions to the public good increase with an increase in the Marginal Per Capita Return and group size (for small MPCRs), but these effects are largely caused by subject confusion rather than altruism or expectations about the minimum profitable coalition.

Decision tasks 1, 2, 3 and 4 and tasks 10, 11, 12 and 13 increase the returns from the public good while holding group size constant and equating internal and external returns from the public good. The corresponding MPCRs within each set of four tasks are: 0.17, 0.33, 0.67 and 1. From Figure 1, consistent with Stylized Fact #2, one can easily see that contributions increase with an increase in MPCR in the human treatment. However, this increase in contributions occurs in both treatments: about 30 tokens from the lowest to highest MPCR. The correlation between the MPCR and contributions is nearly indistinguishable between the treatments: $\rho = 0.97$ in the all-human treatment and $\rho = 0.93$ in the virtual-player treatment.

Turning to group-size effects, note that tasks 1 through 6 are identical to 10 through 15 with the exception that the group size for the latter set is six rather than two. For the human treatment, focusing on tasks with “small” MPCRs (consistent with Stylized Fact #3) – 1, 2 and 5 ($n = 2$) and 10, 11 and 14 ($n = 6$) – average contributions increase from 10.5% to 21.0% of endowment with an increase in group size from two to six. For the virtual treatment, contributions increase from 4.1% to 11.9%. The ratio of the change in virtual-player contributions to the change in all-human contributions suggests that nearly 75% of the group-size effect observed in the all-human treatment is an artifact of confusion.
**Result 3.** Confused subjects use experimental parameters as cues to guide payoff-maximizing contributions, thus confounding attempts to use exogenous changes in parameters to identify other-regarding preferences. After controlling for this confound, we detect weak pure altruism and no warm-glow altruism.

To formally quantify the magnitude of pure altruism and warm-glow, GHL consider theoretical specifications for individual utility and estimate utility function parameters using a logit equilibrium model (see GHL for details). We estimate logit equilibrium models with our data, interpret the estimated parameters, and compare parameters across treatments. Estimated coefficients and standard errors are presented in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>All-Human Treatment</th>
<th>Virtual-Player Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Altruism</td>
<td>Warm-Glow</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.069*</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>$g$</td>
<td>-</td>
<td>1.183*</td>
</tr>
<tr>
<td></td>
<td>(1.183)</td>
<td>(0.300)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>40.846*</td>
<td>35.524*</td>
</tr>
<tr>
<td></td>
<td>(2.205)</td>
<td>(2.433)</td>
</tr>
<tr>
<td>N</td>
<td>720</td>
<td>720</td>
</tr>
</tbody>
</table>

Note: standard errors in parentheses.
* indicates parameter is statistically different from zero at the five percent level.

Test results (all tests allow $\mu$ to vary across treatments).
Altruism model: $\alpha_V = \alpha_H$ $\chi^2(1) = 8.720, p = 0.003$
Warm glow model: $g_V = g_H$ $\chi^2(1) = 8.502, p = 0.004$
Combined model: $\alpha_V = \alpha_H$ and $g_V = g_H$ $\chi^2(2) = 10.068, p = 0.007$

The “altruism” model considers the altruism motive but not warm-glow, the “warm-glow” model considers warm-glow but not altruism, and the “combined” model considers both motives. The parameter $\alpha$ is a measure of pure altruism, the parameter $g$ measures warm-glow,
and $\mu$ is an error parameter. While $\mu$ measures dispersion and does not indicate the magnitude of confusion contributions, GHL argue that statistical significance of this parameter does indicate decision error is present. Focusing first on the all-human treatment, our results mirror those of GHL. In particular, we find that pure altruism and warm-glow altruism parameters are statistically significant (5% level) when considered in isolation, but only the pure altruism parameter is significant in the combined model. Further, estimates of $\mu$ are statistically different from zero for each specification.

Note, however, that the altruism parameter is also positive and statistically different from zero in the virtual-player treatment in which the contributions of “others” are just numbers on pieces of paper in an envelope on the subjects’ desks. There is no altruism in such an environment. Altruism parameters are statistically different across models, but are of similar magnitude. Concentrating on the altruism model, the estimated $\alpha$’s in the two treatments suggests that about half (53.6%) of the altruism detected in the all-human treatment is in fact noise generated by confused subjects using the experimental parameters as cues to guide payoff-maximizing contributions. Our results suggest that the average subject is willing to give up only 3.2 cents to provide a dollar to another person.\footnote{GHL’s estimated $\alpha$ ranges from ten to fourteen cents.} Although this value is statistically different from zero, it is rather small.\footnote{Our results do not stem from our modest changes to the GHL design. We also conducted an exact replication of the GHL instructions with all-human and virtual-player treatments (Cotten, Ferraro and Vossler, forthcoming). The pattern of contributions in relation to design factors is similar to both the GHL results and our current results.}

Based on our post-experiment question about the profit-maximizing contributions level in the virtual-player treatment, at least 28.1 percent of respondents were unable to discern the dominant strategy of zero contributions after participating in this experiment. Using as an upper-bound the percentage who did not free-ride in all rounds except when MPCR=1, as much as
64.6% did not understand incentives. Interestingly, however, nearly all subjects in the all-human treatment, post-experiment question could correctly state the contributions level that would have maximized group earnings. Thus, subjects apparently see that their payoffs increase with increasing group member contributions, but are unable to do one more iterative step of strategic thinking to see that they can benefit from free-riding.

Confused subjects thus use the changes in the parameters across decision tasks as a cue of how to behave. In contrast to the white noise picked up in the logit equilibrium model’s noise parameter, the confusion we identify systematically varies with the parameters designed to identify pure altruism such that some confused individuals “look” like altruists. Based on this finding, one would expect laboratory estimates of altruism to have little external validity. Laury and Taylor (forthcoming) report that subjects who are more ‘altruistic’, based on the GHL experiment design, are less likely to contribute money to an actual tree-planting program, even after controlling for experimental earnings, subject demographics and attitudes. We believe that confusion confounds their comparison.

In the next section, we apply the virtual-player design to the standard repeated-round VCM experiment. We do so to focus on the robust observed dynamics in repeated-round play and on recent claims that conditional cooperation is the main motivation for contributions to the public good rather than confusion.

**IV. EXPERIMENT 2: DYNAMICS OF REPEATED-ROUND VCM GAMES**

Experimental Design

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 MPCR is constant and equal to 0.50, thus making free-riding the dominant strategy and contributing the entire endowment the socially optimal strategy. These attributes of the experiment are common knowledge.

Instructions are presented both orally and in writing (see appendix B). Subjects receive a payoff table that displays 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. 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. The same author moderated all of the sessions.

In the all-human treatment, subjects play 25 rounds of the game. 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 the sessions (subjects separated by dividers).

The virtual-player treatment is identical to the all-human treatment with one exception: each human is aware that he or she is paired with three nonhuman, virtual players and that each virtual player plays a predetermined contribution profile. Subjects are informed that this contribution profile is the same profile produced by a human player in a previous all-human treatment. They are told that a computer 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. As with
all our experiments, subjects are provided these contribution profiles on paper sealed in an
envelope at their desk and reminded 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.

Each session consists of two experimental conditions, with 25 rounds of play in each. We
designate the participants in the first 25-round period in a session as “I” (for “inexperienced”) and
the participants in 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) \textbf{HI}: Participants are \textit{inexperienced}, and play in all-human groups for 25 rounds.

2) \textbf{VI}: Participants are \textit{inexperienced}, and play in virtual-player groups for 25 rounds.

3) \textbf{HE}: Participants are \textit{experienced}, and play in all-human groups for 25 rounds.

4) \textbf{VE}: Participants are \textit{experienced}, and play in virtual-player groups for 25 rounds.

The \textit{HI} condition is the standard linear VCM game about which we wish to draw inferences
about the subjects’ motives. To do so, we contrast \textit{HI} with \textit{VI}, \textit{VE} and \textit{HE}. Subjects in a \textit{VI}
(\textit{VE}) treatment observed the same history of contributions as subjects in a corresponding \textit{HI}
(\textit{HE}) condition: each subject in \textit{HI} (\textit{HE}) has a “twin” in \textit{VI} (\textit{VE}). The only difference between
\textit{HI} (\textit{HE}) and \textit{VI} (\textit{VE}) is that the humans in \textit{VI} (\textit{VE}) were playing with virtual players.
Since HI data are used in the VI treatment and HE data used in the VE treatment, we necessarily ran a sequence of three experiments. HI, by definition, had to be run first. HI subjects played their last 25 rounds in a modified version in which virtual-agent contributions were taken from previous HI, rather than HE, players. These last 25 rounds are thus not used in any analysis. VI subjects participated in VI and then HE. Only after these sessions were complete could the VE sessions, in which subjects first saw the VI contribution profile followed by virtual-agent contributions taken from HE, take place (otherwise any differences among experienced subjects could be due to either playing with virtual players or to seeing a different history of contributions). In sum, HI data come from one experiment, VI and HE data from a second, and VE data from a third experiment.

Two-hundred and forty students from Georgia State University were recruited to participate in a computerized experiment conducted in the laboratory of the Center for Experimental Economics. Eighty were assigned to 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.

Houser and Kurzban’s (2002) (hereafter, HK) design is similar to ours, but there are three important differences. First, aggregate computer contributions to the public good in HK are three-fourths of the average aggregate contribution observed for that round in the human condition. Thus, the identification of contributions attributable to confusion 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 average aggregate contributions from the human
condition thwarts important dynamics. Keser and van Winden (2000), Ashley, Ball and Eckel (2003), and Carpenter (2004) find that contributions are history-dependent.

Second, and related to the role of the history of contributions, HK’s 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.

Third, HK do not attempt to discriminate among different kinds of other-regarding preferences or confusion behaviors, whereas we present in the next subsection a micro-econometric model to undertake this discrimination.9

Microeconometric Behavioral Model

We develop a dynamic model of individual behavior for VCM experiments that encompasses the popular motives for public good contributions discussed in the literature. 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 from previous play (reinforcement learning) or to the contributions from others (herding) as indicators of optimal behavior.

Turning first to reinforcement learners (Roth and Erev, 1995; Mason and Phillips, 1997; Erev and Roth, 1998; Cason and Friedman, 1999), we depict these subjects as engaging in a hill-climbing exercise whereby they search for the profit-maximizing strategy based on financial

---

9 Our sample size is also much larger; HK have only 20 subjects in their all-human sample.
signals from previous rounds. Let $y_{it}$ denote individual $i$’s contribution to the public good in round $t$. Further, let $\pi_{it}$ 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:

$$y_{it} = \beta_{1}^{RL} y_{i,t-1} + \beta_{2}^{RL} y_{i,t-2} + \gamma^{RL} D_{i,t-1} (\pi_{i,t-1} - \pi_{i,t-2})$$  \hspace{1cm} (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 $\gamma^{RL}$ 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:

$$y_{it} = \alpha^{Herd} + \beta_{1}^{Herd} y_{i,t-1} + \beta_{2}^{Herd} y_{i,t-2} + \lambda^{Herd} (y_{i,t-1} - Y_{i,t-1}/n)$$  \hspace{1cm} (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, $\alpha^{H}$, 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 $\alpha_{\text{Herd}} > 0$. Melding (2) and (3) yields our model of virtual-player treatment behavior

$$y_{it} = \alpha_{\text{Herd}} + \beta^V_1 y_{i,t-1} + \beta^V_2 y_{i,t-2} + \lambda^\text{Herd}(y_{i,t-1} - Y_{i,t-1}/n) + \gamma^\text{RL}[D_{i,t-1}(\pi_{i,t-1} - \pi_{i,t-2})] + \epsilon^V_{it}$$  \hspace{1cm} (4)

where $\beta^V_j$ is a weighted average of $\beta^\text{RL}_j$ and $\beta^\text{Herd}_j$ (for $j = 1, 2$), and $\epsilon^V_{it}$ 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 altruism, pure altruism, and conditional cooperation. The standard assumption that warm-glow and pure 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 pure altruism is thus depicted by the relationship between contributions and a constant term. Thus,

$$y_{it} = \alpha_{\text{WG}} + \alpha_{\text{IU}}$$  \hspace{1cm} (5)

where $\alpha_{\text{WG}}$ and $\alpha_{\text{IU}}$ 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 conditional cooperators is:

$$y_{it} = \alpha_{\text{SR}} + \beta^\text{SR}_1 y_{i,t-1} + \beta^\text{SR}_2 y_{i,t-2} + \lambda^\text{SR}(y_{i,t-1} - Y_{i,t-1}/n)$$  \hspace{1cm} (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:

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

where $\varepsilon_{it}^{ORP}$ 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:

$$y_{it} = \alpha^H + \beta_1^H y_{i,t-1} + \beta_2^H y_{i,t-2} + \lambda^H (y_{i,t-1} - Y_{i,t-1}/n) + \gamma^{RL} [D_{i,t-1} (\pi_{i,t-1} - \pi_{i,t-2})] + \varepsilon_{it}^H$$  \hspace{1cm} (8)

where $\beta_j^H$ is a weighted average of $\beta_j^{SR}$ and $\beta_j^V$; $\lambda^H$ is a weighted average of $\lambda^V$ and $\lambda^{SR}$, and $\alpha^H = \alpha^{Herd} + \alpha^{ORP}.$ Contributions data from the all-human treatment alone do not allow one to identify the parameters $\alpha^{Herd}$ and $\alpha^{ORP}$ 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 $\lambda^H$ and $\lambda^C$ is not equal to $\lambda^{ORP}$ – the proportions of conditional cooperators and herders in the sample are not explicitly known – we cannot recover an estimate of $\lambda^{ORP}$ by comparing parameter estimates from comparable all-human and virtual-player treatment data. However, larger estimates of $\lambda$ with all-human treatment data indicate that strong reciprocator behavior is significant.
In estimating the unknown 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 are, by construction, non-negative integers, and there are a preponderance of zeros and small values. The “count data” nature of contributions lends itself well to a Poisson modeling framework. The standard arguments for discrete choice models motivate using a Poisson over OLS: OLS predicts negative values and the discrete nature of the data causes OLS errors to be heteroskedastic. Some recent analyses treat contributions as being censored at both zero and full-endowment and employ two-limit Tobit models (Ashley, Ball and Eckel, 2003; Bardsley and Moffatt, 2003; Carpenter, 2004). However, contributions data are not technically censored, as contributions, in principal, cannot assume negative values and zero values are not due to nonobservability. Maddala (2001, pgs. 335-336) argues that many common applications of the Tobit are inappropriate, and observes that researchers are tempted to apply the Tobit whenever many observations of the dependent variable take on a zero value.

We account for unobserved, individual-specific heterogeneity (i.e., the panel structure of the data) by using a Poisson quasi-MLE. Specifically, this is the Poisson MLE coupled with White’s (1982) robust covariance estimator, a.k.a. the “sandwich” estimator, adjusted for clustering at the individual level. This estimator is robust (i.e., consistent) to a variety of misspecifications, including distributional misspecification. Note that while we argue in favor of using the Poisson, our qualitative results are consistent with those obtained from OLS. For comparison purposes we also estimated two-limit Tobit models with random-effects, but the models did not converge when applied to data from three of our four experimental conditions.

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. We estimated models (available upon request) that included up to five-period lags. Inferences drawn from these alternative specifications are similar to those presented below.\textsuperscript{10}

Results

Figure 2 presents average contributions by round as a percentage of endowment for each of the four experimental conditions. Table 2 presents the estimated coefficients from the Poisson behavioral model corresponding to each experimental condition.

Result 4. \textit{Other-regarding preferences and confusion are significant motives that determine public good contributions in the standard repeated-round VCM experiment (HI). Furthermore, other-regarding behavior contributions decrease over rounds.}

Comparing all-human and virtual-player contribution rates with inexperienced subjects represents the cleanest distinction between contributions stemming from other-regarding motives versus those due to confusion in the standard VCM game. Mean contributions to the public good in HI, 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).

\begin{table}[h]
\centering
\caption{Dynamic Poisson Models of Individual Behavior}
\begin{tabular}{lcccc}
\hline
Dependent variable is $y_{it}$ (\textit{i}'s contribution to the public good in round \textit{t}) & All-Human, & Virtual-Player, & All-Human, & Virtual-Player, \\
\hline
\end{tabular}
\end{table}

\textsuperscript{10} In a similar vein, we investigated more general econometric specifications that allowed the slopes on the “feedback” and group behavior variables to depend on whether the deviations were positive or negative. We failed to reject our more parsimonious specification using conventional tests.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>inexperienced</th>
<th>inexperienced</th>
<th>experienced</th>
<th>experienced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\alpha$</td>
<td>1.8516</td>
<td>1.2482</td>
<td>1.3803</td>
<td>0.9379</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0738)**</td>
<td>(0.1353)**</td>
<td>(0.1001)**</td>
<td>(0.1235)**</td>
</tr>
<tr>
<td>$y_{i,t-1}$</td>
<td>$\beta_1$</td>
<td>0.0337</td>
<td>0.0376</td>
<td>0.0497</td>
<td>0.0462</td>
</tr>
<tr>
<td>[subject contributions in round $t-1$]</td>
<td></td>
<td>(0.0028)**</td>
<td>(0.0047)**</td>
<td>(0.0037)**</td>
<td>(0.0059)**</td>
</tr>
<tr>
<td>$y_{i,t-2}$</td>
<td>$\beta_2$</td>
<td>0.0141</td>
<td>0.0298</td>
<td>0.0150</td>
<td>0.0353</td>
</tr>
<tr>
<td>[subject contributions in round $t-2$]</td>
<td></td>
<td>(0.0016)**</td>
<td>(0.0034)**</td>
<td>(0.0024)**</td>
<td>(0.0038)**</td>
</tr>
<tr>
<td>$y_{i,t-1} - (Y_{i,t-1}/n)$</td>
<td>$\lambda$</td>
<td>-0.0153</td>
<td>-0.0072</td>
<td>-0.0256</td>
<td>-0.0143</td>
</tr>
<tr>
<td>[deviation from average contributions of other group members in round $t-1$]</td>
<td></td>
<td>(0.0023)**</td>
<td>(0.0040)*</td>
<td>(0.0035)**</td>
<td>(0.0039)**</td>
</tr>
<tr>
<td>$D_{i,t} (\pi_{i,t-1} - \pi_{i,t-2})$</td>
<td>$\gamma$</td>
<td>0.0044</td>
<td>0.0062</td>
<td>-0.0003</td>
<td>0.0039</td>
</tr>
<tr>
<td>[profit “feedback” mechanism]</td>
<td></td>
<td>(0.0018)**</td>
<td>(0.0033)*</td>
<td>(0.0037)</td>
<td>(0.0045)</td>
</tr>
<tr>
<td>Log-Likelihood</td>
<td></td>
<td>-13,751.38</td>
<td>-12,538.29</td>
<td>-13,530.07</td>
<td>-10,134.32</td>
</tr>
<tr>
<td>$N$</td>
<td></td>
<td>1840</td>
<td>1840</td>
<td>1840</td>
<td>1840</td>
</tr>
</tbody>
</table>

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.

In comparison, VI contributions start at 28.5% and fall to 9.8% by round 25. On average, subjects contribute 32.5% and 16.8% of all endowments to the public good in the all-human and virtual-player treatments, respectively. Dividing VI contributions by HI contributions suggests that 51.6% of the total contributions in the standard VCM game stem from confusion; the remaining 48.4% are attributable to other-regarding behavior. Statistical tests (Kolmogorov-Smirnov) indicate that public good contributions are statistically different, and higher, in the all-human treatment at the 5% level both on average and in 24 of 25 rounds (see appendix D).

In their related study, HK 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 HK experiment, our figure is 53%. These summary statistics are quite close. However, note that
HK 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). In contrast, our rate of decline is statistically different and is about 1.8 times *faster* for the all-human treatment, indicating that other-regarding behavior declines over rounds.\(^\text{11}\)

Ledyard (1995, p.146) conjectures that confused subjects might simply split their endowment approximately half-half to see what happens. Our first-round data support his conjecture: in HI, 31 subjects chose a contribution between 20 and 30 tokens and in VI, 29 subjects chose a contribution between 20 and 30 tokens. Note that in HI, 11 subjects contributed their entire endowment, while none did in the VI, suggesting that most of the full-endowment contributors are not confused.

**Result 5.** *Result 4 is robust to experience.*

Although experienced subjects in the all-human treatment contribute less than inexperienced subjects (HI vs. HE), a finding consistent with the literature, the relationships observed between virtual-player and all-human treatments with inexperienced subjects are robust to experience (HE vs. VE). That is, there is statistical evidence that contributions stemming from other-regarding behavior are significant and are *decreasing* over rounds. In particular, other-regarding behavior is significant and decreases over rounds.

---

\(^{11}\) 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 HI and VI conditions are: [HI] contributions = 49.05 – 1.27*round; [VI] contributions = 25.84 – 0.69*round. A likelihood ratio test rejects the hypothesis of equal slope coefficients for two experiment conditions \(\chi^2(1)=13.43, p<0.01\).
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.\(^\text{12}\)

**Result 6.** *Contribution rates are similar across inexperienced and experienced subjects in the virtual-player treatment. Thus there is no evidence that increasing awareness of the dominant strategy among confused players drives the decay of contributions over time.*

If a substantial component of the decay in contributions across rounds stems from subjects becoming aware of (i.e., “learning”) the dominant strategy of zero contributions (Stylized Fact #1), the contributions from inexperienced subjects in VI should be significantly higher than contributions from experienced subjects in VE. The data do not support this prediction. Average contributions are 16.8% and 11.9% of endowment with inexperienced and experienced subjects, respectively. Using a Kolmogorov-Smirnov Test where the average contribution across rounds from an individual is used as an independent observation, 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 (see appendix D).

While this pattern suggests that a few inexperienced subjects may have indeed (quickly) learned the dominant strategy, overall learning effects appear to be minimal. Analysis of the numbers of free-riders ($0 contribution) by round yields a similar conclusion: in Round 1, Round 2, Round 24 and Round 25 of VI, there were 22, 27, 49 and 50 free-riders; the corresponding numbers in VE are 38, 31, 46 and 58.

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.

\(^{12}\) Using the framework outlined in footnote 8, the estimated relationships for the HE and VE conditions are: [HE] contributions = 31.26 – 0.79*round; [VE] contributions = 18.19 – 0.48*round. A likelihood ratio test rejects the hypothesis of equal slope coefficients for these experiment conditions [\(\chi^2(1)=4.63, p=0.03\)].
(which reflect behavior in past all-human sessions). We test this alternative hypothesis directly through our econometric model.

**Result 7.** *The majority of the decline in contributions in the virtual-player treatment with inexperienced or experienced subjects arises from herding behavior.*

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 significance 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 fail to reject the hypothesis that contributions due to reinforcement learning are statistically different from zero in periods 9-25 (we fail to 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.

Additional supporting evidence for Result 7 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 (for these players: mean = 28 tokens; median = 25 tokens).\(^{13}\) 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 HI. Our

\(^{13}\) 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.
focus group results (next subsection) suggest the proportion is much higher and also indicates that many of the subjects we identify as herders actually believe they are playing an Assurance Game.

**Result 8.** *Conditional cooperation is a motive for contributions in the all-human treatment.*

For both experienced and inexperienced subjects, the estimate of $\lambda$ 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, conditional cooperation is a motive for contributions in the all-human conditions ($\lambda^H > \lambda^{Herd}$).

Results 7 and 8 thus imply that 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. Conditional cooperators 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 group interactions in order to make valid inferences.

**Result 9.** *Little warm-glow or pure altruism is evident in the all-human treatment.*

Finally, given the standard assumption that warm-glow and pure altruism 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/pure altruism contributions. For inexperienced subjects, we have that the average subject contributes 4.23% of their endowment due to warm-glow and pure altruism. For experienced subjects, this figure is 2.3%. Putting this into another perspective,
at most just 13.0% and 11.0% of observed contributions across rounds could be attributed to
warm-glow and pure altruism for inexperienced and experienced subjects, respectively.

Thus, in the absence of punishment opportunities, the co-existence of free riders, conditional cooperators 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 conditional cooperators and the herding on the downward trend by confused players.

Our results imply that much of the contributions observed in VCM experiments come from confused individuals who never 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 (VE). If we base our estimate on behavior in rounds 10-19 in HI and VI (after hill-climbing has been abandoned), we estimate that about half of subjects are confused (53%), about a fifth are free-riders, and a quarter are conditional cooperators.

Focus Group

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 written and oral feedback to the experimenters. These subjects had just completed playing 50 rounds with virtual players. Everyone stated in writing and then orally that they were
playing with nonhuman agents with pre-determined decisions that could not be affected by the actions of the humans.

The short answer to our question is “Yes, many subjects are oblivious to the dilemma inherent in the public goods game induced in the laboratory.” More detailed results are contained in the appendix, but the bottom-line is that approximately 1 in 3 subjects correctly understood the incentives. The remainder either found the incentives undecipherable and herded on virtual player contributions (sometimes after unsuccessful hill-climbing exercise) or erroneously believed they were playing an Assurance Game. The latter is a coordination game with two pure-strategy equilibria, one of which is Pareto optimal, and the other which is inefficient but is less risky as there is less payoff variance. In the public goods game setting, Assurance Game players do not construe zero contributions as free-riding, but as the risk-dominant strategy. For subjects playing an Assurance Game, play is not affected by the presence of non-human virtual players. Assurance Game subjects become angry with group members (human or virtual) because they fail to coordinate on more lucrative equilibria, not because they are free-riding.

V. EXPERIMENT 3: CONFUSION MIMICS CONDITIONAL COOPERATION

Experimental Design

Fischbacher, Gächter and Fehr (2001) (hereafter, FGF) use a variant of the strategy-method in a one-shot VCM (p.397) “that directly elicits subjects’ willingness for conditional cooperation.” Their innovation is the use of a “contribution table,” which asks subjects to consider possible average contribution levels of the other group members, and state how much they would contribute to the public good conditional on each level. Each subject must indicate an
unconditional contribution (traditional VCM design) and fill out a contribution table. Within each group of four players, the total contributions to the public good are determined by the unconditional contribution of three players and the relevant conditional response in the fourth player’s contribution table.

Importantly, FGF claim (p.398) that having subjects “answer 10 control questions that tested their understanding of this public good problem…indicates that the subjects understood the mechanics and the implications of the above payoff function.” In other words, FGF believe their subjects are not confused about the incentives in the game. FGF classify 50% of the subjects as true conditional cooperators, 14% as “hump-shaped” contributors who conditionally cooperate up to about half of endowment and then decline, 30% as free riders, and the rest as “other patterns.” Other studies using the same design (Keser and van Winden, 2000; Fishbacher, Gächter, and Fehr, 2001; Fischbacher and Gächter, 2004; Burlando and Guala, 2005; Houser and Kurzban, 2005; Chaudhuri, Graziano and Maitra, 2006; Chaudhuri and Paichayontvijit, 2006) find higher rates of true conditional cooperators compared to free riders, but essentially similar frequencies of these player “types.”14

To test whether the addition of control questions mitigates confusion, and in a related vein whether the FGF cleanly measures conditional cooperation, we ran 80 additional subjects from Georgia State University through an experiment identical to FGF with the exception that participants were matched with virtual players. We use the same set of control questions FGF

---

14 A recent study using the design in a repeated-round context (Fischbacher and Gächter, 2004) claims the vast majority of contributions are motivated by conditional cooperators with no evidence of pure or warm-glow altruism. They claim confusion accounts for few contributions to the public good (“at most 17.5 percent,” p.3). They also argue that most of the decay in contributions is not from learning but from the interaction among free riders and conditional cooperators who revoke their cooperation once they realize they are among people who are not “norm abiders.”
claim ensures proper understanding of the game’s incentives. The instructions are included as appendix C.

Results

The results are strikingly similar to FGF (see p.400) despite the fact that our subjects were not grouped with human beings. Using FGF’s criteria for classifying subjects, we are forced to classify 53% of our sample as conditional cooperators, 23% as free-riders, 15% as hump-shaped contributors, and the rest as “other patterns.” It appears that the pre-experiment test questions do little to mitigate confusion.

**Result 10.** Recent experiments employing the strategy method erroneously classify many confused individuals as conditional cooperators.

Herders and Assurance Game players behave in similar fashion to conditional cooperators – if the experimenter asks them how much they would contribute if the other group members invested X tokens, one would likely see a high correlation between subject answers and X. The presence of herders and Assurance Game players suggests why recent studies on conditional cooperation find little confusion and a lot of conditional cooperation. The confused subjects of previous studies (Andreoni, 1995; Palfrey and Prisbey, 1997; Houser and Kurzban, 2002; Ferraro, Rondeau and Poe, 2003) have been reclassified as conditional cooperators, “hump-shaped” contributors and “others.”

---

15 The same problem is inherent in designs that identify conditional cooperation through a positive correlation between own contributions and beliefs about the contributions of others.
VI. CONCLUSION

Decision errors, confusion, noisy behavior are familiar concepts in experimental economics. However, as suggested by Hey (2005, p.325): “…the source and possible nature of the noise are rarely explicitly discussed…. If one makes the wrong assumptions about the … noise, then one usually makes wrong inferences from the data.” Although past public goods experiments have identified “confusion” – behavior that stems from subjects’ inability to discern the nature of the game in which they are playing – as a substantial source of contributing to the public good, past experimental work either ignores this confusion or treats it as no more than (random) statistical noise. Continuing in this fashion is perilous.

We use novel experimental designs, combined with microeconometric behavioral models that place structure on the decisions of confused players, to identify confusion behavior. We demonstrate that the stylized facts of twenty-five years of public good experiments may be artifacts of the instructions and designs commonly used by experimentalists. In the web of confusion, subjects use experimental parameters and (in repeated games) the behavior of other subjects as behavioral cues. These confusion behaviors manifest themselves as other-regarding behaviors and thus serve to confound analysis by, for example, distorting the effects of increasing the MPCR or estimates of altruism. This confusion does not disappear with repetition, as suggested by some, nor is it picked up in the noise parameters of quantal response models. Coinciding with the notion that context beyond the control of the experimenter can affect the interpretation of results (see Levitt and List, 2006), we find that at least some of the confusion arises because participants instead see themselves as playing an Assurance Game.

We are not saying the stylized facts don’t exist. After netting out confusion contributions with our virtual-player design, we still find evidence of pure altruism and conditional
cooperation, for example. However, given that we cannot draw inferences for about half our subject pool it is not clear that qualitative results such as MPCR effects or decline in contributions in repeated-round games will prevail in the absence of rampant confusion.

Field experiments may be less exposed to problems associated with confusion. However, we believe that economists should not abandon lab experiments. Instead, as Levitt and List (2006) argue, we should anticipate the types of biases common to the lab, and design experiments to minimize such biases. Why is there so much confusion? From our experiments we hypothesize that confusion stems from the use of: (1) instructions that are stripped of familiar contextual cues; and (2) incomplete payoff tables. By framing the contribution decision as one of “investing” between a “private account” and a “group account”, some of our focus group participants thought they were playing an Assurance Game and, when asked what they thought the experiment was about, many subjects said “investment in the stock market.” Further, standard instructions only provide information on returns from the public good for different levels of aggregate contributions. Additional computations are needed for a player to determine that, holding the contributions of others’ constant, individual earnings are maximized by contributing nothing.

The next crucial step in this line of research is the development of instructions to mitigate confusion through the use of “associative framing” and/or the provision of more complete payoff information. While these designs were not implemented to mitigate confusion per se, a handful of public goods experiments and related experiments have investigated the effect of using associative framing or more complete payoff information and find very strong treatment effects (Saijo and Nakamura, 1995; Cookson, 2000; Charness, Freschette and Kagel, 2004; Rege and Telle, 2004). Evidence from two recent pilot experiments imply that these instruction changes
also reduce confusion, suggesting that the strong treatment effects observed stem from
systematic differences in contributions decisions when the majority of participants are not
confused. In Cotten, Ferraro and Vossler (forthcoming) we report results from a classroom pilot
eperiment in which we used associative framing. When decisions are framed as “contributions
towards the provision of a public good”, 22 of 25 subjects in a 10-round VCM game are able to
identify the dominant strategy before the experiment begins. In a second pilot experiment, we
provided a payoff matrix that contained information on own and average other group-member
earnings, conditional on own and average contributions from other group members. Except for
the augmented payoff table, the experiment was similar to Experiment 2, with 25 rounds of the
all-human treatment followed by 5 rounds of the virtual-player treatment. We estimate that just
one-quarter of contributions in the first 5 rounds of the all-human treatment stem from confusion,
and 19 of 20 were able to identify the dominant strategy following the experiment.16

Given the simplicity of the VCM game environment and the ubiquity of abstract
instructions without familiar contextual cues, we believe our results have important implications
for the burgeoning use of laboratory methods to test economic theories. When using abstract
instructions it is likely that subjects will introduce their own context in order to make palatable
the decision task. When subjects are confused about incentives and the parameters are changing
as they make decisions, they infer that the parameter changes must be a signal that their
decisions ought to be changing. Experiments using within-subject designs are more likely to
experience this confound. In conclusion, we believe that an important area for future research in
experimental economics will be to identify the source and nature of the noise in experimental

16 Given the within-subject design and associated spillover effects from playing with humans, we purport that this
estimate of confusion contributions is an upper bound.
games and to develop ways to reduce this noise when it appears to be an artifact of the experimental design rather than part of the decision process being studied.

References


Figure 1. Experiment 1, Mean Contributions per Decision Task by Experiment Condition

<table>
<thead>
<tr>
<th>Decision Task</th>
<th>0%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

$n$:

| AL-Human treatment | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   |
| Virtual-Player treatment | 6   | 6   | 6   | 6   | 6   | 6   | 6   | 6   |

$m_l$:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>4</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

$m_c$:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>4</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>6</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>
Figure 2. Experiment 2, Mean Contributions per Round by Experiment Condition