Experimental Economics, 5:39–51 (2002)

Learning to Punish: Experimental Evidence from a Sequential Step-Level Public Goods Game

DAVID J. COOPER
Department of Economics, Weatherhead School of Management, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
email: djc13@po.cwru.edu

CAROL KRAKER STOCKMAN
Graduate School of Public Health, University of Pittsburgh, 130 DeSoto Street, A648, Pittsburgh, PA 15261, USA
email: stockman@pitt.edu

Abstract

This paper studies how subjects in a three-person sequential step-level public good game learn to punish free riders more over time. Our current work makes several additions to the literature on other regarding behavior. First, our experiment provides evidence that subjects care about the actions that lead to an outcome as well as the outcome itself, replicating the results of A. Falk, E. Fehr and U. Fischbacher (*Economic Inquiry*, in press), J. Brandts and C. Solà (*Games and Economic Behavior*, 36(2), 138–157, 2001.) and J.H. Kagel and K. Wolfe (Working paper, Ohio State University, 1999). Second, our experiment provides one of the first tests of the newer theories of reciprocity by A. Falk and U. Fischbacher (Working paper, University of Zurich, 2000) and G. Charness and M. Rabin (*Quarterly Journal of Economics*, in press) that take a psychological games approach. We find that these theories fail to explain the experimental data. Finally, we examine the mechanism by which subjects learn to punish free-riding more ofter over time.

Keywords: public goods, learning, reciprocity

JEL Classification: H4, C7, C9

1. Introduction

One of economics' most successful collaborations between theory and experiments has taken place in the study of "other-regarding behavior." This term is a catchall for a wide variety of anomalous behavior observed in experiments that is not consistent with modeling subjects as pure payoff maximizers, but instead suggests that subjects have preferences over the payoffs and (possibly) actions of other players.[1] Recent work by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) has shown that relatively simple models that allow players' preferences to depend on the payoffs of others as well as their own payoffs can explain a wide variety of anomalous experimental results.

While these two papers are a tremendous improvement on their predecessors, they share common weaknesses. In both papers, players' preferences are fixed and depend solely on the realized outcome of the game. Subsequent experimental papers have raised problems with both of these assumptions. Cooper and Stockman (in press) find evidence in a public goods experiment that subjects' preferences are not fixed, but instead must be learned. Studies of cardinal ultimatum games by Falk et al. (in press) and Brandts and Solà (2001) and of three-person ultimatum games by Kagel and Wolfe (1999) suggest that subjects care about the actions that lead to payoffs rather than just caring about the payoffs alone. Responding to the latter set of experimental results, papers by Falk and Fischbacher (2000) and Charness and Rabin (in press) present models of other-regarding behavior in which preferences depend upon the actions and intents of other players as well as the final outcome.

In this paper we study a three-person sequential step-level public good game in which, contrary to standard theoretical predictions, subjects learn to punish free riders more often over time. Our current work makes several additions to the literature on other-regarding behavior. First, our experiment provides evidence that subjects care about the actions that lead to an outcome as well as the outcome itself, replicating the results of Falk et al., Brandts and Solà, and Kagel and Wolfe. Second, our experiment provides one of the first tests of the newer theories of reciprocity by Falk and Fischbacher and Charness and Rabin that take a psychological games approach. We find that these theories fail to explain the experimental data. Finally, we examine the mechanism by which subjects learn to punish free-riding more often over time. Econometric analysis reveals clear patterns underlying changes in subjects' choices. In the conclusion, we discuss what sort of models of learning and/or other-regarding preferences would be consistent with these patterns.

2. The MCS game, the experimental design, and predicted results

We study a 3-person step-level public goods game known as the Minimal Contributing Set (MCS) game.[2] The game tree for the MCS game is shown in figure 1. Each player is given an identical initial endowment of tokens sufficient to ensure that negative payoffs are not possible. Players sequentially choose whether or not to contribute some of these tokens to provision of a public good. Contribution is a binary choice; players choose to either contribute or not contribute, and a predetermined cost of contribution is deducted from a player's initial endowment following contribution. The public good is provided if at least 2 of the 3 players contribute. If the public good is provided, then each of the players receives an additional 18 tokens regardless of whether he contributed to the public good. Costs of contribution are deducted regardless of whether the public good is provided. Players are given perfect information—they know all players' payoff tables and the moves of preceding players.

For both treatments, Player 1's cost of contribution is 1 token and Player 3's cost of contribution is 16 tokens. In the low Player 2 cost treatment, Player 2's cost of contribution is 3 tokens. In the high Player 2 cost treatment, Player 2's cost of contribution is 14 tokens. Thus, Player 3 faces an identical decision in either treatment if Player 1 has contributed and Player 2 has not. Only one cost structure is used in any given experimental session. Although all players in both treatments are better off contributing and having the public good provided than not contributing and not having the public good provided (contribution costs are less than 18 tokens for all players), each player earns more if he can free-ride on the contributions of the other players.

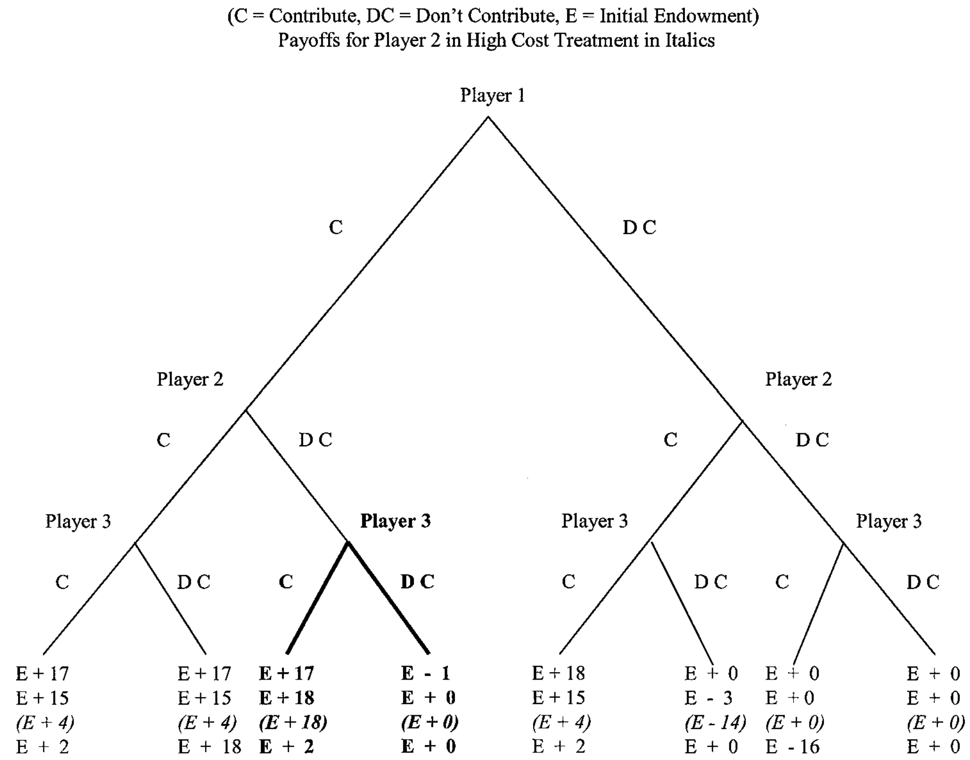

Figure 1. The MCS game.

Equilibrium analysis makes a good starting point for predicting play in the MCS game. Restricting our attention to pure strategy Nash equilibria, and assuming that utility is a function solely of a player's monetary payoff, the MCS game has four Nash equilibria. In one equilibrium none of the players contribute and the public good is not provided. In the remaining three equilibria exactly two of the three players contribute and the public good is provided efficiently. Only the equilibrium in which the last two players contribute is a subgame perfect equilibrium.

The structure of the game suggests ex ante that fairness considerations will push play away from the subgame perfect equilibrium. Consider the position of a "critical" third player, a third player who knows that one of the preceding players has contributed while the other has not. A critical third player faces no strategic uncertainty, and therefore knows that the public good will be provided if and only if he contributes. If this player cares solely about his monetary payoffs, he should always contribute. However, this player also knows that a preceding player with a lower cost of contribution has not contributed and will free-ride on provision of the public good. We can easily imagine that a critical third player who cared about relative payoffs might resent this, and therefore not contribute to punish the earlier player who did not contribute. Such actions serve to destabilize the subgame perfect outcome.

Our analytical focus will be on the behavior of critical third players when the first player contributed and the second player did not. The resulting subgame is highlighted in bold type on figure 1. (Unless otherwise noted we will always be referring to this particular case when we refer to critical third players.) We focus on this specific subset of the data for two reasons. First, we want to see how attitudes towards fairness evolve for a choice where no strategic uncertainty exists to confound matters. Second, we want the distribution of monetary payoffs over the three players that the third player is choosing between to be identical in the two treatments. This allows us to focus on the psychological games aspect of critical third players' behavior. We now turn to predicting how the two treatments will differ for this subset of critical third players.

The leading models of interpersonal preferences, as presented in Bolton and Ockenfels and Fehr and Schmidt, allow preferences to depend solely on realized payoffs. Specifically, when critical third players are choosing following a contribution by Player 1, only the payoff distributions that could possibly result should affect their decisions. The possible outcomes that were eliminated when Player 2 choose not to contribute should have no impact on Player 3s' decisions. Therefore, either model predicts identical behavior over the two treatments for critical third players.

There exists ample experimental evidence to suggest that behavior does not depend solely on realized outcomes, but also on foregone outcomes and the perceived intentions of players. Recognizing this, Falk and Fischbacher and Charness and Rabin include these psychological games factors into subjects' preferences. Either model predicts that the contribution rate by critical third players will be higher in the high Player 2 cost treatment than in the low Player 2 cost treatment. Intuitively, the presumable reason a critical third player chooses not to contribute is resentment that Player 2 is free-riding. It makes sense that this resentment should be less intense when the cost of contribution was higher for Player 2. Thus, we expect a higher contribution rate in the high Player 2 cost treatment.[3]

The preceding models are static. However, Cooper and Stockman find that the decisions of critical third players over identical payoff distributions change over time, often in a direction that leads to lower monetary payoffs. We expect a similar dynamic to operate in our current experiment. We therefore expect that treatment effects will only emerge gradually over time as subjects' behavior evolves.

To summarize, we make three initial predictions. (1) Contribution rates by critical third players will differ between the two treatments. (2) These differences will grow over time. (3) The contribution rate will be higher in the high Player 2 cost treatment than in the low Player 2 cost treatment.

3. Experimental procedures

One hundred and seventy-seven undergraduate students from the University of Pittsburgh and Carnegie Mellon University were recruited to participate in this experiment. Four sessions of each treatment were run, with 84 subjects in the high Player 2 cost treatment and 93 subjects in the low Player 2 cost treatment.[4] A minimum of fifteen and a maximum of thirty subjects were used in each session. Average payoff was $18 for a one-and-a-half hour session.

Before the beginning of a session, instructions were read aloud to all subjects. Subjects were also given a written copy of the instructions and of the payoff tables. All substantive questions about the instructions were answered out loud to ensure common knowledge. Subjects were asked to complete a short payoff quiz to verify their ability to read the payoff tables.

Each session consisted of forty periods of play of the MCS Game. Subjects were told the number of periods to be played. We randomly determined each subject's player position (Player 1, Player 2, or Player 3) before the first round of play. A subjects' position remained unchanged throughout the course of the session. Subjects knew their own position and that there were an equal number of subjects in each position, but did not know the position of any other subject. At the beginning of each round of play, we randomly and anonymously assigned subjects to three person groups containing one subject for each position. The instructions emphasized that the three person groups would be randomly rematched in every round. This was intended to preserve the one-shot strategic character of the MCS game. After each round of play, subjects were informed of the decisions made and the payoffs earned by the other members of their group. Subjects were given record sheets on which to record their outcomes. While we did not require subjects to fill out the record sheet, we observed that most did.

To avoid any framing effects, neutral terms were used wherever possible. For example, players were asked to choose between "X" and "Y" rather than "contribute" and "don't contribute." No explicit mention was ever made of "provision" or "public goods". Instead, the payoff table refers solely to "your choice" and "total number of other players choosing X."

For the low Player 2 cost treatment, each player received an initial endowment of 12 tokens. To avoid the possibility of negative payoffs for Player 2s and any resulting loss aversion, we increased the initial endowment in the high Player 2 cost treatment to 16 tokens. The show-up payment was lowered to keep total payouts constant.

Subjects were paid privately, in cash, based on their token earnings in two periods randomly chosen from the forty periods of play. Each token was worth \$0.30 to subjects.

4. Results

The individual level data are summarized in Table 1.[5] Before turning to the choices of critical third players following a contribution by Player 1, a few notes about the play of the preceding two players. First, the contribution rate for Player 1s is somewhat higher for the first twenty periods of the high Player 2 cost treatment, albeit not by a statistically significant amount. This difference vanishes in later periods. For both treatments, the contribution rate by Player 1 rises steadily over time, reflecting the relatively low probability that the other two players will provide the public good if Player 1 does not. Regardless of the Player 1's choice, the contribution rate for Player 2s is much lower throughout in the high Player 2 cost treatment, presumably due to the large difference in costs. In both cases, the diference between treatments grows over time.

We now focus on the choices of critical third players contingent on a contribution by Player 1. Figure 2 displays the evolution of critical third players' contribution rates for the two treatments, broken down into five period segments. For both treatments, critical third

Table 1. Summary of contribution rates, high player 2 cost treatment vs. low player 2 cost treatment contingent on role and preceding players' actions.

	Player 1		Player 2 Pl. 1 contributed		Player 2 Pl. 1 didn't contribute		Critical player 3 Pl. 1 contributed		Critical player 3 Pl. 2 contributed	
	High cost	Low cost	High cost	Low cost	High cost	Low cost	High cost	Low cost	High cost	Low cost
Periods 1–10	.721 202/280	.700 217/310	.193 39/202	.465 101/217	.410 32/78	.699 65/93	.632 103/163	.638 74/116	.625 20/32	.892 58/65
Periods 11–20	.771 216/280	.742 230/310	.162 35/216	.422 97/230	.375 24/64	.775 62/80	.536 97/181	.564 75/133	.458 11/24	.581 36/62
Periods 21–30	.779 218/280	.787 244/310	.206 45/218	.504 123/244	.403 25/62	.864 57/66	.434 75/173	.504 61/121	.640 16/25	.491 28/57
Periods 31–40	.818 229/280	.819 254/310	.205 47/229	.591 150/254	.275 14/51	.786 44/56	.429 78/182	.625 65/104	.286 4/14	.522 23/44

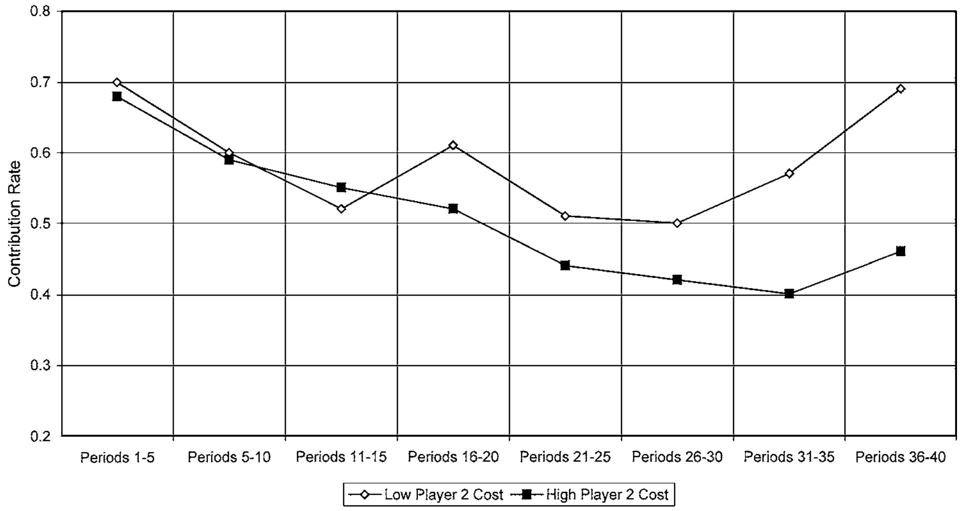

Figure 2. Comparison of contribution rates by critical third players between the low and high player 2 cost treatments.

players' contribution rates trend down over time but hook up at the very end. Critical third players' contribution rates start at about the same place for both treatments, but begin to diverge in the second half of the experiment. By the end of the experiment, the contribution rate for critical third players is substantially lower in the high Player 2 cost treatment than the low Player 2 cost treatment. The difference is 19% over the last ten periods and 23% over the last five periods. The raw data clearly support our prediction that the critical third players' contribution rates will differ in the two treatments, and that the difference will grow over time. However, the sign of the effect is not what we predicted. Rather than having a higher contribution rate, the high Player 2 treatment has a lower contribution rate for critical third players. This result is inconsistent with the predictions of either the Falk and Fischbacher or Charness and Rabin models.

One notable feature of the raw data, especially in the low Player 2 cost treatment, is a hook upwards in the contribution rates for critical third players over the last ten periods. At first glance, this late increase suggests that some sort of reputation building by third players may be occurring. In fact, what this late increase illustrates is an upwards bias introduced into the raw data by aggregation effects.

In later periods of the experiment, the probability of being a critical third player is endogenous. If learning by the first two players moves them towards strategies with higher expected payoffs, there will be positive correlation between the initial probability that critical third players contribute in a session and the probability that third players are critical in later periods. It follows that observations of critical third players in later periods are more likely to come from sessions in which critical third players tend to contribute. This aggregation effect biases estimated changes in the contribution rate for critical third players upwards with the bias increasing over time.

Aggregation effects cause the late increase in critical third players' contribution rates for both treatments. When the data is broken down by session, there in fact does not exist any clear upwards trend over time. For four of the eight sessions, the contribution rate by critical third players actually drops between periods 21–30 and periods 31–40. Weighing all eight sessions equally, the average increase between periods 21–30 and periods 31–40 is only 3.4%, with averages of 0.5% for the four high Player 2 cost sessions and 6.3% for the four low Player 2 cost sessions. The upward trends shown in figure 2 are driven by an increasingly large proportion of observations from the sessions with high contribution rates for critical third players, not a change in the individual level behavior of Player 3s.

We now turn to formal econometric analysis of the data. This analysis serves two purposes. First, it allows us confirm that the treatment effect we observe with the naked eye is statistically significant. In particular, the formal econometrics allows us to control for the strong individual effects (and resulting aggregation bias) in the data. Second, the econometrics provides us with some insight into the forces underlying the observed treatment effect.

Table 2 reports the results of two probit regressions on the data. The data set consists of all observations of a critical third player's decision following contribution by a Player 1. To control for individual effects (and the resulting aggregation effects), we use a random effects specification.[6] For both models the dependent variable is the critical third player's decision ($0 =$ don't contribute, $1 =$ contribute). Thus, positive (negative) parameter estimates correspond to higher (lower) contribution rates by critical third players. The random effects term is significant at the 1% level in both regressions, but is not reported in Table 2 since it is of no direct interest.

Model 1 measures the statistical significance of the treatment effect we observe in the raw data. Right hand side variables for this model are a dummy for periods 21–40 and an interaction term between the dummy for periods 21–40 and a dummy for the high Player 2 cost treatment.[7] The dummy for the second half of the experiment is negative and significant

Table 2. Probit regressions with a random effects specification, critical third players, player 1 contributed, 59 subjects and 1173 observations.

Variable	Model 1	Model 2
Constant	1.061** (.076)	3.064** (.294)
Periods 21–40	−.376** (.085)	2.000** (.552)
High player 2 cost × periods 21–40	−.347** (.094)	−.158 (.219)
Pl. 1 contribution rate × periods 1–20		−3.515** (.336)
Pl. 1 contribution rate × periods 21–40		−6.264** (.770)
Pl. 2 contribution rate (conditional on Pl contributing) × periods 1–20		.814** (.238)
Pl. 2 contribution rate (conditional on Pl contributing) × periods 21–40		.079 (.419)
Pl. 2 contribution rate (conditional on Pl not contributing) × periods 1–20		−.107 (.197)
Pl. 2 contribution rate (conditional on Pl not contributing) × periods 21–40		−.241 (.285)
Log likelihood	−518.81	−489.48

+Significantly different from 0 at the 10% level.
*Significantly different from 0 at the 5% level.
**Significantly different from 0 at the 1% level.

at the 1% level. Regardless of treatment, contribution rates for critical third players are dropping significantly over time. The estimated parameter for the treatment effect in the second half of the experiment is also negative and statistically significant at the 1% level. The estimated marginal effect of the high cost treatment for periods 21–40 is −10.3%. This is somewhat smaller than the −12.9% difference observed in the raw data over these periods, presumably reflecting the effect of aggregation bias. Model 1 confirms that the observed treatment effect is statistically significant even after controlling for individual effects.

Given that the treatment effect only develops gradually and that subjects observe quite different histories of play by the other two players across the two treatments, it seems natural to focus on what Player 3s might be learning that differs between the two treatments. Model 2 adds several variables to Model 1 that measure the past behavior by Player 1s and Player 2s that the Player 3 in question has observed. This regression lets us determine what aspects of others' past behavior have explanatory power for the current behavior of Player 3s.

Model 2 adds three measures of other players' past behavior to Model 1. For each observation we calculate the average contribution rate the subject has observed for Player 1s in earlier periods and the average contribution rate the subject has observed for Player 2s conditional on Player 1's decision in earlier periods.[8] These variables are interacted with dummies for periods 1–20 and periods 21–40, giving us a total of six new variables. The results are striking. Player 1s' past contribution rate has an enormous negative impact on the current contribution rate by critical third players. This effect is statistically significant at the 1% level in both halves of the experiment and becomes significantly stronger over time. For the final twenty periods, a 10% increase in the past contribution rate by Player 1s has an estimated marginal effect of decreasing the contribution rate for critical third players by 22.7%. In contrast, the contribution rates for Player 2 matter little. For the first twenty periods, the contribution rate by Player 2s contingent on Player 1 contributing is statistically significant at the 1% level, but the marginal effect is less than a quarter of the marginal effect from Player 1s' contribution rate. Over the final twenty periods, when the treatment effect actually emerges, Player 2s' contribution rate contingent on Player 1 contributing virtually equals zero and no longer has a statistically significant effect. The contribution rate for Player 2s, contingent on Player 1 *not* contributing, never has a statistically significant effect. The parameter estimate for the treatment effect in Model 2 is less than half of the estimate for Model 1, and it fails to achieve statistical significance at any standard level. Over half of the treatment effect appears to be driven by the relatively small difference between the two treatments in contributions by Player 1s. Even though the difference between the treatments in contribution rates by Player 2s is enormous, this plays little role in generating the treatment effect.[9]

The effect of the contribution rate for Player 1s in past periods is sufficiently strong that it can be seen with the naked eye. For each subject who was a Player 3, we calculate the number of times in the first 10 periods the subject was matched with a Player 1 that contributed. We also calculate each subject's contribution rate over the final twenty periods as a critical third player. We then group the subjects into cells by how many Player 1 contributions they saw in the first ten periods, and average the contribution rates for the final twenty periods across the subjects in each cell. Figure 3 graphs the resulting data. The bars are labeled with the number of subjects for each cell. The negative relationship between the two variables

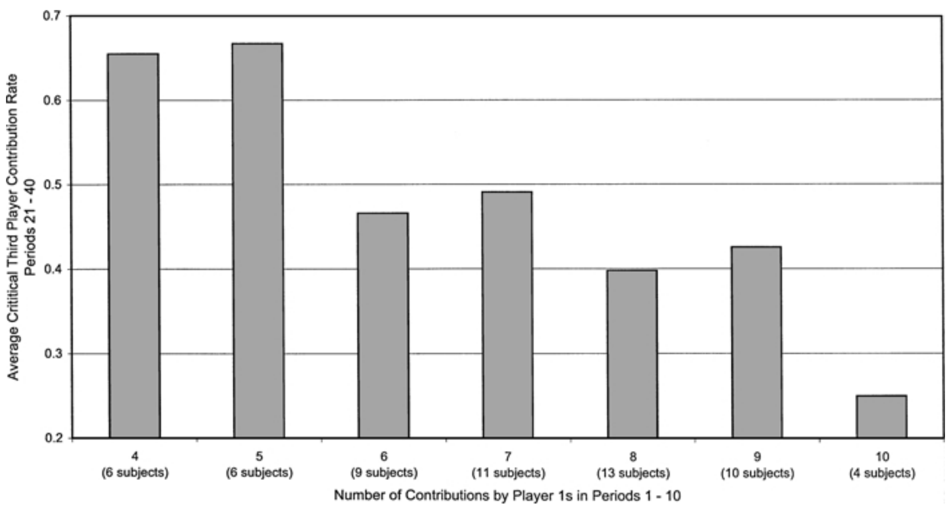

Figure 3. Effect of past player 1 contributions on critical third player contributions.

is obvious. A third player that initially observes a high number of contributions by Player 1s is far less likely in later periods to contribute when critical following a contribution by a Player 1.

To summarize, the data yield clear results. The high Player 2 cost treatment yields a significantly lower contribution rate for critical Player 3s than the low Player 2 cost treatment. This effect only emerges in the second half of the experiment. Looking at features of the past history of play that drive the behavior of critical third players, a dominant role is played by the past contribution rate of Player 1s. The past contribution rate of Player 2s, regardless of Player 1's action, has little impact on the current behavior of critical third players.

5. Discussion

This paper makes several contributions to the existing literature on other regarding behavior. Our data allows us to cleanly reject both the leading models of other-regarding preferences (Bolton and Ockenfels and Fehr and Schmidt) as well as newer models incorporating aspects of psychological games (Falk and Fischbacher and Charness and Rabin). Moreover, our data also yields strong results about the forces underlying the failure of these models' predictions. Subjects' current actions depend strongly on the actions by others that they have observed earlier in the experiment.

In thinking about why the Falk and Fischbacher and Charness and Rabin models fail to predict the observed treatment effect, consider what features of the models underlie their predictions. For both models, in weighing how badly a Player 2 has behaved by not contributing, a player should consider how costly a contribution would have been. In other words, Player 3 should consider whether contributing by Player 2 was a reasonable sacrifice. Given that the predicted differences between the treatments don't emerge even in the early periods, it appears that this hypothesized sensitivity to others' costs is weaker than the theories would suppose. This point may simply be too subtle for subjects to grasp. More generally, these models are failing not so much because of what is in them but because of what isn't in them. The strong dynamic aspect of third players' choices that is captured by the econometrics in Model 2 is lacking from these static models.

The specific pattern of critical third players' reactions to the past behavior of Player 1s and Player 2s is difficult to explain. Since the treatment effect only emerges gradually over time, explanations that relate to subjects' learning are natural candidates. One possible explanation uses the hybrid model of learning and fairness presented in Cooper and Stockman. This model combines reinforcement learning in the spirit of Roth and Erev (1995) with Bolton and Ockenfels' model of other-regarding preferences. In this framework, contribution rates for critical third players drop as these individuals learn that they are better off (it terms of their utility) not contributing than allowing an earlier player to free ride. The speed of this learning should depend on how often subjects are getting experience as a critical third player—more experience leads to faster learning. Since subjects in the high Player 2 cost treatment get more experience as a critical third player, we would predict lower contribution rates by critical third players in this treatment, as is observed.

A second explanation relies on subjects who are learning social norms. Briefly, a subject's perceptions of the behavior of other players may depend on what he considers to be normal behavior for those other players. For example, if virtually all Player 1s contribute to the public good, such a contribution would not be considered an unusually kind act. This implies that subjects' behavior will change as they learn about the behavior of others. Given that the experience subjects receive in the two treatments is quite different, the resulting differences in learned norms can lead to a treatment effect. While this hypothesis does not make a clear directional prediction ex ante, given the observed differences in behavior by Player 1s and Player 2s between the two treatments, we would predict a lower contribution for critical third players in the high Player 2 cost treatment as observed.

The working paper version of this manuscript includes a detailed discussion of these hypotheses. Briefly, while both of these learning hypotheses predict the main treatment effect, both miss other features of the data (in particular, the lack of response to past play by Player 2s). A theory that combines these two approaches fits the data well, but has the drawback of being quite complex. Moreover, since the experiments reported here weren't designed to directly test these hypotheses, we cannot conclusively accept or reject either one. Further experiments are needed to definitively establish a cause for the unexpected treatment effect we find.

It should also be noted that while we have focused on the dynamic aspects of the treatment effect, there is a residual treatment effect remaining (albeit not a statistically significant one) after we control for the past behavior of Player 1s and Player 2s. If this effect turns out to be robust in further experiments, it would be interesting in its own rights. A non-learning explanation for the treatment effect that we are interested in exploring is social distance. There is extensive evidence from the psychology literature that individuals tend to compare themselves more with people who they view as being in a similar situation (e.g., Hoffman et al., 1954). For example, as an assembly line worker I compare my well-being with the well-being of other assembly line workers rather than the well-being of the firm's CEO. One possible effect of increasing Player 2's cost of contribution is to make Player 2's situation more similar to that of Player 3. This in turn might make Player 3 more upset if Player 2 is doing much better than him by free riding, leading to the observed treatment effect of lower contribution rates by critical third players. While this hypothesis is purely speculative at this point in time, it is one that we plan to examine in future experiments.

Acknowledgments

The authors would like to thank the National Science Foundation for financial support. We would like to thank Eric Bettinger, Jordi Brandts, Gary Charness, Sue Helper, John Kagel, George Loewenstein, Jim Rebitzer, Al Roth, and Bob Slonim for their helpful comments. We would also like to thank the editors of Experimental Economics, Charlie Holt and Arthur Schram, and an anonymous referee for their help in preparing this manuscript. The usual caveat applies.

Notes

1. Well known examples include the ultimatum game (Roth, 1995), the dictator game (Roth, 1995), VCM public goods games without thresholds (Ledyard, 1995), experimental labor markets (Fehr et al., 1998a; Fehr et al., 1998b), and trust games (Berg et al., 1995).

2. The MCS game was first introduced in a simultaneous form by Van de Kragt et al. (1983) and in a sequential form by Erev and Rapoport (1990).
3. For an outline of how these predictions are made formally, see the working paper version of this manuscript.
4. Data from the low Player 2 treatment has been previously reported in Cooper and Stockman as the 1/3/16 treatment in that paper.
5. Table 1 does not report results for third players who are not critical. There are 864 such observations, 637 in which both Player 1 and 2 contributed and 227 in which neither Player 1 nor Player 2 contributed. In none of these 864 observations did the third player contribute.
6. The random effects specification controls for correlation between observations from the same individual but not for correlation between observations from different individuals in the same session. While Player 3s never directly interact, such correlations could develop through their interactions with Player 1s and Player 2s. Model 2 directly examines the effects of past experience with Player 1s and Player 2s on Player 3s' choices.
7. We use this specification to control for changes over time because of its simplicity. Regressions with a variety of other specifications reach qualitatively similar conclusions. Of particular interest are regressions in which the data is broken up into ten period blocks. Regardless of whether these time dummies are interacted with a treatment dummy or not, we find no evidence of a statistically significant increase in contribution rates from periods 21–30 to periods 31–40. This provides additional evidence that the late increase in contribution rates observed in the data is due to aggregation effects, not reputation effects. As an alternative to Model 1, we ran a regression that also included a dummy for the high Player 2 cost treatment without any interaction. This specification tests for treatment effects in the first twenty periods. Consistent with the raw data, the additional variable fails to achieve statistical significance at any standard level.
8. The use of lagged variables raises a problem for observations in which the lagged variable can't be defined. We have set the lagged variable equal to its average for these observations.
9. Averaging over all 580 observations in our data set for periods 21–40, the average value of Player 1's historical contribution rate is 71.4% in the low Player 2 cost treatment and 76.5% in the high Player 2 cost treatment. Naively, multiplying the marginal effect by the difference in treatments gives us a treatment effect of 10.7% through Player 1's contributions. We have rerun model 2 with the past contribution rates by Player 1s and Player 2s interacted with a treatment dummy for the final 20 periods. None of these three additional parameters are statistically significant at even the 10% level, nor do the three parameters achieve joint significance ($x^2 = 2.4$, 3 $d.f.$, $p > .10$). This result suggests that critical third players respond to the past behavior of Player 1s and Player 2s in the same fashion for both of the treatments.

References

Berg, J., Dickhaut, J.W., and McCabe, K.A. (1995). "Trust, Reciprocity and Social History." *Games and Economic Behavior*. 10(1), 122–142.

Bolton, G.E. and Ockenfels A. (2000). "ERC: A Theory of Equity, Reciprocity and Competition." *American Economic Review*. 90(1), 166–193.

Brandts, J. and Solà, C. (2001). "Reference Points and Negative Reciprocity in Simple Sequential Games." *Games and Economic Behavior*. 36(2), 138–157.

Charness, G. and Rabin M. "Social Preferences: Some Simple Tests and a New Model." *Quarterly Journal of Economics*, in press.

Cooper, D.J. and Stockman, C.K. "Fairness and Learning: An Experimental Examination." *Games and Economic Behavior*, in press.

Erev, I. and Rapoport, A. (1990). "Provision of Step-Level Public Goods: The Sequential Contribution Mechanism." *Journal of Conflict Resolution*. 34(3), 401–425.

Falk, A. and Fischbacher U. (2000). "A Theory of Reciprocity." Working paper, University of Zurich.

Falk, A., Fehr, E., and Fischbacher, U. "On the Nature of Fair Behavior." *Economic Inquiry*, in press.

Fehr, E., Kirchler, E., Weichbold, A., and Gachter, S. (1998a). "When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets." *Journal of Labor Economics*. 16(2), 324–351.

Fehr, E., Kirchsteiger, G., and Reidl, A. (1998b). "Gift Exchange and Reciprocity in Competitive Experimental Markets." *European Economic Review*. 42, 1–34.

Fehr, E. and Schmidt, K.M. (1999). "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*. 114, 817–868.

Hoffman, P.J., Festinger, L., and Lawrence, D.H. (1954). "Tendencies Toward Group Comparability in Competitive Bargaining." *Human Relations*. 7, 141–160.

Kagel, J.H. and Wolfe, K. (1999). "Testing Between Alternative Models of Fairness: A New Three-Person Ultimatum Game." Working paper, Ohio State University.

Ledyard, J.O. (1995). "Public Goods: A Survey of Experimental Research." In J.H. Kagel and A.E. Roth (eds.), *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

Roth, A.E. (1995). "Bargaining Experiments." In J.H. Kagel and A.E. Roth (eds.), *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

Roth, A.E. and Erev, I. (1995). "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economic Behavior*. 8(1), 164–212.

Van de Kragt, A.J., Orbell, J.M., and Dawes, R.M. (1983). "The Minimal Contributing Set as a Solution to Public Goods Problems." *American Political Science Review*. 77(1), 112–122.