JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2004, **82,** 161–176 NUMBER 2 (SEPTEMBER)

CONTINGENCIES OF REINFORCEMENT IN A FIVE-PERSON PRISONER'S DILEMMA

RICHARD YI AND HOWARD RACHLIN

STATE UNIVERSITY OF NEW YORK AT STONY BROOK

As in studies of self-control, a tit-for-tat contingency in an iterated prisoner's dilemma game creates a conflict between maximization of local and global reinforcement. The present experiments examine this conflict in a multiplayer prisoner's dilemma game. Versus tit for tat, cooperation corresponds to self-control; defection, always immediately reinforced, corresponds to impulsiveness. Three experiments examined sensitivity of behavior to the global reinforcement contingency imposed by tit for tat. Undergraduates played a five-player prisoner's dilemma game against four dummy players programmed to play tit for tat or randomly. With tit for tat, a player's cooperation (or defection) increased dummy players' cooperation (or defection) on subsequent trials—reinforcing cooperation and punishing defection in the long run. Participants cooperated at a higher rate when the dummy players played tit for tat than when the dummy players played randomly. These results are consistent with findings in corresponding studies of self-control. Some participants, caught in a trap of mutual defection with the tit-for-tat playing dummy players, came to cooperate when the tit-for-tat contingency was reset ("forgiving" participants' previous defections) during a pause in the game. This increase was a result of the combined effects of a pause and reset; neither pausing nor resetting independently resulted in an increase in cooperation.

Key words: prisoner's dilemma, self-control, reinforcement, tit for tat, choice, humans

---

The matrix of Figure 1 is an example of available choices and rewards in a traditional two-player prisoner's dilemma game (PDG). Each player has two choice alternatives: "cooperation" and "defection." Consider Player A's alternatives: If B cooperates, then A earns 10 units for defection or seven units for cooperation; if B defects, then A earns three units for defection or zero units for cooperation. Regardless of B's choice, A's reward is higher for defection than for cooperation. Player A should defect. Because the game is perfectly symmetrical, the same is true for Player B. However, when both players defect, both earn less (three units each) than if both players had cooperated (seven units each). This conflict between individual and group benefits creates the dilemma.

If the game is iterated over many trials (IPDG), players may apply a tit-for-tat strategy (TFT) to facilitate mutual cooperation (see Komorita, 1976 for a review). A player applying TFT mimics the other player's previous-trial choice: after cooperating on the first trial, cooperation is reciprocated with cooperation on the next trial or defection with defection on the next trial. Suppose Player B plays TFT. Against B's TFT strategy, A's present-trial choice is mimicked by B's next-trial choice. Note that when B cooperates, A's alternatives are higher (10 or seven units) than when B defects (three or zero units). Thus, when B plays TFT, A's cooperation on the present trial is reinforced (with a delay) by B's cooperation on the next trial, and A's defection on the current trial is punished (with a delay) by B's defection on the next trial (Rachlin, Brown, & Baker, 2001). Note also that in an IPDG against TFT, overall reinforcement rate is directly proportional to percentage cooperation. With the values in Figure 1, and one player playing TFT, the other player's reinforcement rate is maximized (at seven units per trial) by 100% cooperation and minimized (at three units per trial) by 100% defection.

Most laboratory studies of social cooperation use a two-player PDG like that of Figure 1. Recently, in order to better mirror real-life social dilemmas, investigators have expanded the number of players in the PDG to more than two (NPDG; Ledyard, 1995). Figure 2 shows the contingencies, for a given player, in a five-player NPDG. The two solid diagonal lines show the reward amounts for cooperation or defection (ordinate) as a function of the number of other players cooperating

The authors would like to thank Ralph Molaro and Glenn Hudson for their programming expertise, and Xochitl de la Piedad and Federico Sanabria for their assistance with data collection. The research was supported by a grant from the National Institutes of Health.

Address correspondence to Richard Yi, Department of Psychiatry, University of Vermont, Burlington, Vermont 05401-1419 (e-mail: ryi@uvm.edu).

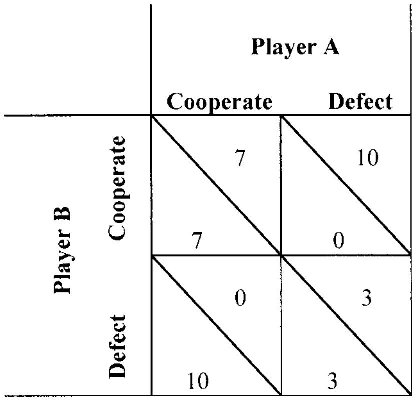

Fig. 1. The choices and outcomes of a prisoner's dilemma game. Each player has two alternatives (cooperation and defection); for each player, the reward for defection (10 units if other player cooperates, three units if the other player defects) is always greater than the reward for cooperation (seven units if the other player cooperates, zero units if the other player defects) on a single trial.

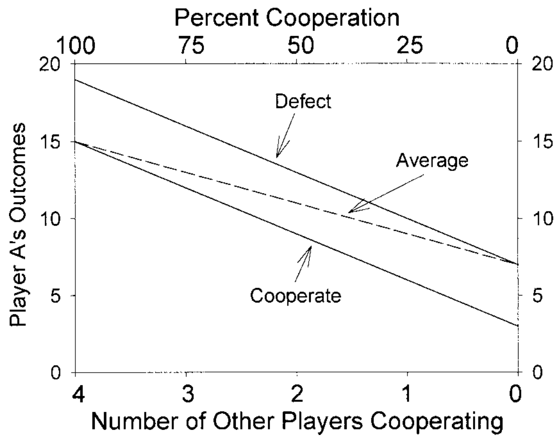

Fig. 2. The contingencies for a single player in a five-player prisoner's dilemma game. The reward for defection (upper solid line) is always greater than the reward for cooperation (lower solid line). The outcome on a given choice is determined by the player's choice and the number of other players who cooperate (*x*-axis). The dashed line is the mean reward for all players as a function of the number of players that cooperate. This mean reward is higher on the left side of the figure, when all other players cooperate.

(lower abscissa). As in the two-player game, the immediate reward for defection is always greater than that for cooperation. In Figure 2, the difference between current-trial defection and cooperation rewards is four units regardless of the number of other players cooperating (the solid lines are parallel). However, the fewer the number of cooperators, the lower the reward for *both* cooperation and defection (the solid lines slope downward as the number of players cooperating decreases). The dashed line shows the average reward over all participants as a function of the number of cooperators. Although defection pays more than cooperation for any individual player, cooperation pays more than defection for the group as a whole. As with the two-player PDG, the dilemma is one of individual versus group benefit.

The present experiments studied the effect of TFT on cooperation in an iterated NPDG (INPDG) with human participants. The TFT contingency was imposed as follows: Participants were told that they were playing a repeated game with 4 other participants (being simultaneously tested) but each was actually playing against 4 dummy players (DPs) whose choices were determined by a computer. If the participant cooperated on any given trial, 1 more DP would cooperate on the next trial; if the participant defected on any given trial, 1 more DP would defect on the next trial (within the limits of 0 to 4 DPs cooperating). In Figure 2, cooperation on a given trial would move the outcomes of a player's alternatives one unit to the left on the next trial (increasing value); defection would move the outcomes of a player's alternatives one unit to the right (decreasing value). Overall reinforcement rate with this INPDG versus TFT (as with the two-player IPDG vs. TFT) is proportional to percentage cooperation over repeated trials; the dashed line in Figure 2 gives overall average reinforcement per trial as a function of percentage cooperation (upper abscissa).[1]

The reinforcement contingencies in an IPDG against TFT correspond to those in

[1] Another method used to program reciprocity with NPDG (Komorita, Parks, & Hulbert, 1992) fixes the number of other players reciprocating the participant's choices. However, this method has two problems. First, all TFT players must play identically (as one): the more other players who play TFT, the lower the effective number of players, and if all other players played TFT, the *N*-player game would become essentially a two-player game. Second, this method leaves open what strategy the other players, not playing TFT, should employ. Any systematic play by these other players may affect the participant's choices. Even random play (to which the best response is repeated defection) has its effect—as the present experiments will show.

studies of self-control. In fact, Harris and Madden (2002) directly compared measures of self-controlled/impulsive behavior (Mazur's [1987] discounting parameter obtained from delay discounting procedures) and behavior in a two-player IPDG against a TFT strategy. They found a modest positive correlation between rate of delay discounting and rate of defection. Except for the DP context, the current procedure is similar to self-control experiments (Ainslie, 1992; Rachlin, 2000) and particularly to ''primrose path'' self-control procedures (Herrnstein, Loewenstein, Prelec, & Vaughan, 1993; Heyman, 2003; Kudadjie-Gyamfi & Rachlin, 1996, 2002). Participants in self-control studies typically choose between a smaller-sooner reward (SS) and a larger-later reward (LL). Preference for the LL reward (and the resulting higher overall reinforcement rate) is said to be ''self-controlled,'' whereas preference for SS (and the resulting lower overall reinforcement rate) is said to be ''impulsive.'' To obtain the higher reinforcement rate in a self-control study, the participant must reject a reward available immediately or after a brief delay for a larger reward available only after a longer delay. Correspondingly, to obtain a higher reinforcement rate in IPDG against TFT, a player must repeatedly ''cooperate''—choose the lower-valued of the current-trial alternatives (seven over 10 units in Figure 1); otherwise, on the next trial, the player would face a pair of still lower-valued units (zero vs. three units in Figure 1).

In the primrose path experiments noted above, participants were told (correctly) that they were playing against a computer and that their reward depended on their own behavior. The primrose path procedure studied what Rachlin (2000) calls ''complex ambivalence,'' a choice between maximizing local reinforcement (corresponding to defection in an INPDG vs. TFT) and maximizing overall, or global, reinforcement (corresponding to cooperation in an INPDG vs. TFT). Participants in primrose path experiments typically did not maximize overall reinforcement rate (nor did they minimize it). Rather, their behavior was found to be responsive to variables that typically influence self-control: the reward for impulsive behavior (as represented by the vertical distance between the solid lines in Figure 2); the reward for self-control (as represented by the slope of the dashed line in Figure 2); the degree to which choices are restructured into larger units either by extrinsic signals or by patterns in the choices themselves (Kudadjie-Gyamfi & Rachlin, 1996). The purpose of the present experiments was to study this primrose path self-control procedure in a context of social cooperation.

GENERAL METHOD

Apparatus

Six networked IBM®-compatible PCs were used, with five serving as the interface for each of the 5 participants in a session (client machines) and one serving as the data collecting (server) machine. Client machines had 14 in. to 16 in. (35.6 cm to 40.6 cm) color monitors and were located in individual cubicles formed by 5 ft (152.4 cm) tall partitions such that other machines and participants were not visible when a participant was seated and facing the monitor. The server machine was located in a different room. It was used to set the experimental conditions and monitor participant choices during the session. The experiment was programmed in C++ using Borland's® C++ Builder 5.0.

One postexperiment questionnaire was administered in Experiment 1. In this questionnaire, participants reported the number of people in the session that they knew, if/how it affected their choices, their strategy during the session, if they would change their strategy if given another opportunity, and their degree of belief in the instructions (to determine knowledge about the deception). Two questionnaires, distracter and postexperiment, were each used in Experiments 2 and 3; these questionnaires together addressed the same points as the one questionnaire of Experiment 1. The distracter task required participants to report if they knew any of the other participants in the room and if/how this affected their choices to that point. The postexperiment questionnaire asked participants to report their strategies in each phase, if they would change their strategy, and degree of belief in the instructions. The primary purpose of these questionnaires was to determine if participants knew of or became aware of the deception.

General Procedure

After participants were seated at their computers, they were asked to read the directions on the monitor. The directions, which initially appeared in a white dialogue box on the center of the screen, were:

> Thank you for agreeing to participate in this experiment. For each trial of this experiment, you will be asked to select between two choices, ⟨X⟩ and ⟨Y⟩. You will earn money for each choice, with the amount depending on your choice as well as the choices of the four other participants.
>
> The payoff will follow these rules:
>
> X: (N × 3 cents) + 7 cents
> Y: (N × 3 cents)
>
> N = # of people choosing Y.
>
> If you have any questions, please ask the experimenter before beginning the experiment. Remember, no talking is allowed.
>
> Please do not begin the experiment until instructed to do so.

After participants reported having read the directions on the monitor, the experimenter elaborated on the instructions. Participants were reassured that they would earn credit for participation in the experiment as well as earn a sum of money. The payoff equations for choices X and Y were repeated to them, and examples of potential earnings on a trial were given. Figure 2 represents the game and possible outcomes on each trial as a function of other players' choices, with the upper and lower lines representing the payoffs for defection and cooperation, respectively. Participants were told that choice X would earn 4¢ more than choice Y on any given trial (the 7¢ bonus minus 3¢ due to reduction of N by 1). A hypothetical trial in which 3 participants chose X and 2 chose Y was used to illustrate this point. Participants were also told that the group as a whole would earn more when relatively more participants selected Y than X. Hypothetical trials in which all participants chose X and all participants chose Y were used to illustrate this point. Participants were informed that they would be given feedback in the white dialogue box in the center of the screen after each trial after all participants had made their choice. Participants were asked to attend to this feedback because it would vary depending on what they chose as well as what other players chose. Participants were asked to remain in their seats, facing their monitors, for the duration of the experiment. Any questions were answered and the session began when all participants said that they understood what was being asked of them. They were not informed of the number of trials per session, but were told that the entire experiment was expected to take about 45 min.

Participants began the experiment by using the computer mouse to click on a button labeled <BEGIN> below the dialogue box. At this point, the directions disappeared from the dialogue box, and the phrase "please make your choice now" appeared. The payoff equations moved from the dialogue box to the left side of the monitor, where they remained for the duration of the session. Two buttons, labeled <X> and <Y>, simultaneously appeared below the dialogue box, with the <X> button left of center and the <Y> button right of center. After the player clicked one of the two buttons, both buttons disappeared and the phrase "please wait for others to make their choices" appeared in the dialogue box. Feedback appeared in the dialogue box 5 s after all players had made their choices. This facilitated the deception that they were playing with the others in the room. Feedback was present on the monitor for 15 s, and the next trial followed immediately. This feedback included the participant's choice and earnings on the previous trial, the number of others who chose each alternative and their respective earnings on the previous trial, and the participant's total earnings. It looked as follows:

> You chose ___ and made ___ cents. ___ other(s) chose X and made ___ cents each. ___ other(s) chose Y and made ___ cents each.
>
> Total earnings = ___

After the appropriate number of trials specified by the experimental condition, the phrase "experiment is complete" appeared on the screen. Participants were then given a postexperimental questionnaire to determine the degree to which they knew the other participants. Participants were also asked what strategy they applied and what the ideal strategy was. Strategies participants reported applying in all experiments reported here were generally straightforward, such as ''pick X (or Y) all the time.'' Additionally, they were asked if they believed the experimenter's explanation of the payoffs. No significant relation was found between any of the task measures and expressed friendship with other players tested at the same time. Data obtained from any participant who reported either knowing ahead of time or figuring out the deception over the course of the experiment would not have been included in the analysis. However, none of the participants reported knowing about the deception involved in the study. In Experiments 2 and 3, the distracter questionnaire used between phases asked only about the relation between participants.

Primary Experimental Manipulation

Participants, although believing they were playing the INPDG against the 4 other participants in the room as stated in the directions, actually played against 4 DPs. Choices by the DPs were determined by the computer in one of two ways:

Tit-for-tat contingency (TFT). The computer played a modified form of TFT in which the distribution of DPs cooperating/defecting reciprocated the participant's choices in the previous trials. The distribution of DP choices is represented on the abscissa of Figure 2, moving from right to left as more DPs cooperate. In this variation of TFT, a player's cooperation on trial N was followed by cooperation by 1 more DP on trial N + 1, and would be reflected by a move from right to left in Figure 2. A defection on trial N was followed by a defection by 1 more DP on trial N + 1, and would be reflected by a move from left to right in Figure 2. The limit on the number of DPs that could cooperate or defect on any given trial was equal to the total number of DPs: 4. If a player were to cooperate on four consecutive trials, all 4 DPs would cooperate on the fifth trial. If a player were to defect on four consecutive trials, all 4 DPs would defect on the fifth trial. All 4 DPs cooperated on the first trial. This followed Axelrod's (1984) suggestion to be ''nice'' (i.e., not to be the first to defect) so as to initiate cooperative behavior.

Random (RAN). The distribution of DPs who cooperated/defected was randomly determined by the computer from trial to trial. The number of DPs cooperating on each trial (1, 2, 3, or 4) was randomly picked by the computer. The remaining DPs defected. In this condition, a participant's choices had no influence on the future choices of the DPs.

Preliminary Experiment

In a preliminary experiment, 30 participants (six groups of 5 players each) played the actual INPDG with each other for 20 trials. That is, there were no DPs. For each trial, the value of N was truly determined by the number of players who cooperated; this value, together with each player's choice to cooperate or defect, determined payoffs. A very low cooperation rate was obtained (mean of all participants = 0.20 cooperations per trial; median of group averages = 0.18), similar to results from other INPDG and social dilemma studies using five or more players at once. In this condition, the behavior of each player had an impact on the rewards obtained by all other players; thus individual participant observations were not independent—each five-player group must be treated as one. Analysis of the 6 five-player groups in this experiment suggests that players in different groups often had vastly different reinforcement patterns in the INPDG. Although the members of four of the five-player groups cooperated at a rate near the overall mean, members of two groups did not: for one, mean cooperation rate was 0.42; for the other, mean cooperation rate was 0.05.

EXPERIMENT 1

METHOD

Participants

Sixty participants (20 male) were recruited from the Psychology Department participant pool at the University at Stony Brook. Participants earned one class credit and also were paid the amount of money they earned over the course of the experiment. Five participants were tested together per session.

Procedure

Thirty participants played the INPDG in the TFT condition for 50 trials, with the other 30 playing in the RAN condition. Each experimental session was conducted with 5 participants to facilitate the deception, but data obtained from these participants were inde-

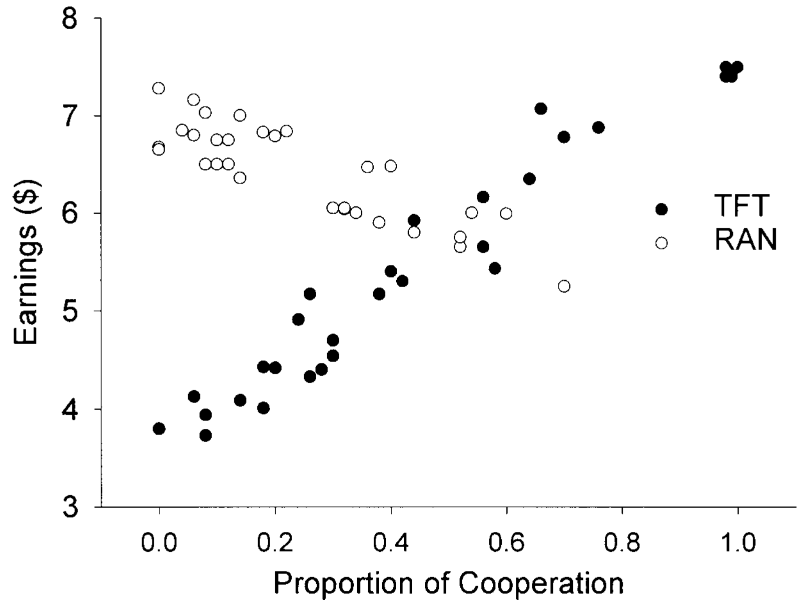

Fig. 3. A scatter plot of the total money earned and proportion of cooperation for each participant in the TFT (filled circles) and RAN (open circles) conditions. The positive slope from TFT and negative slope from RAN indicate that obtained reinforcement rates were consistent with the programmed reinforcement rates.

pendent. Feedback appeared in the dialogue box 5 s after the last participant made his or her choice, and was presented in the dialogue box for 15 s before the next trial began. Participants were directed to fill out the post-experiment questionnaire following the last trial. Participants were then debriefed, paid, and dismissed.

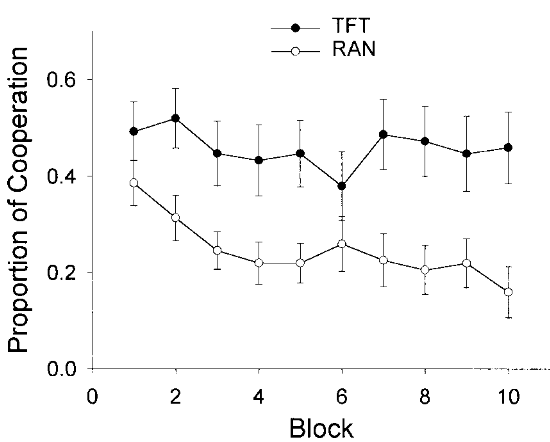

Fig. 4. The mean cooperation rate in 10 five-trial blocks versus both TFT (filled circles) and RAN (open circles) of Experiment 1. Error bars indicate standard error of the mean.

Results

As reported earlier, none of the participants reported suspicion of deception in this study. No systematic patterns of results were observed with questionnaire data and will not be reported here. Figure 3 is a scatter plot of the amount of money earned against TFT and RAN as a function of the proportion of cooperation. The full range of cooperative behavior was observed in TFT, from complete defection to complete cooperation. In RAN, no individual continued cooperating unreinforced, and mostly noncooperative behavior was observed. The near-linear positive slope in TFT and the near-linear negative slope of RAN verify that the obtained reinforcement rates are consistent with the programmed reinforcement rates.

Trials in both conditions were divided into 10 five-trial blocks; analyses were conducted on mean cooperation rate for each block for each condition (Figure 4). Cooperation rates were similar across conditions at the beginning of the session, with mean cooperation slightly higher in the TFT condition than in RAN. As the session progressed, cooperation rate remained fairly constant in TFT, whereas it quickly declined in RAN.

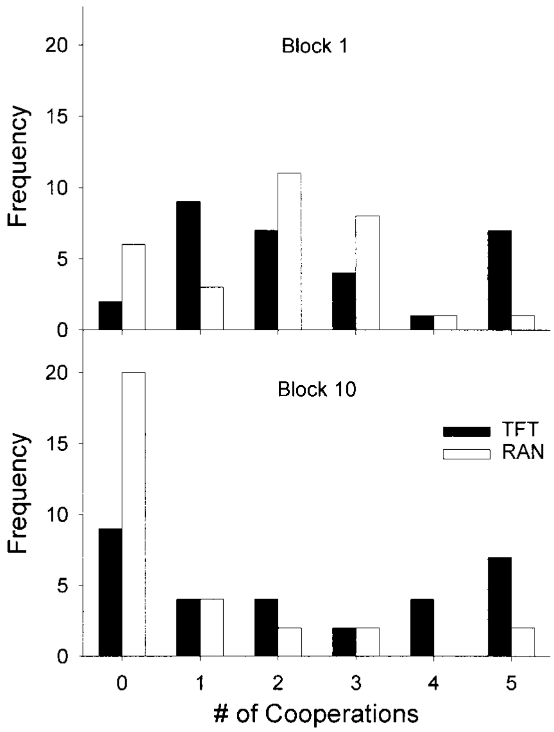

Fig. 5. The number of participants as a function of number of cooperations in Blocks 1 and 10 versus both TFT (filled bars) and RAN (open bars) of Experiment 1.

A 2 × 10 mixed analysis of variance (ANOVA) was conducted on these effects. Cooperation rate was significantly higher versus TFT ($M = 0.46$) than RAN ($M = 0.25$), F (1, 58) = 10.00, $p < .05$. A planned comparison was conducted between mean proportions of cooperation in Block 1 for TFT ($M = 0.49$) and RAN ($M = 0.39$). As expected, this difference was not significant, F (1, 58) = 1.91, $p > .05$. However, because the difference in means was moderate, the first choice of all players was examined to confirm that players in the two groups had similar expectancies or experiences prior to beginning the session. Of the 30 players in the TFT group, 15 cooperated on the first trial. Of the 30 players in the RAN group, 13 cooperated on the first trial. An additional planned comparison revealed a significant difference between mean proportions of cooperation in Block 10 for TFT ($M = 0.46$) and RAN ($M = 0.16$), F (1, 58) = 10.93, $p < .05$.

Figure 5 shows frequency distributions of the number of participants and number of cooperations in Blocks 1 and 10. In Block 1, most players cooperated at least once: of the 30 players, 28 and 24 cooperated at least once in the TFT and RAN conditions, respectively. In Block 10, however, a majority of players (20 of 30) in the RAN condition exclusively defected, whereas only 2 of 30 players exclusively cooperated: The distribution is positively skewed. Many players in the TFT condition also defected exclusively (9 of 30) in Block 10, but there were also nearly as many (7 of 30) who cooperated exclusively. This distribution is bimodal, with the major mode at exclusive defection and the minor mode at exclusive cooperation.

The top graphs of Figure 6 show cumulative counts of the cooperation choices for some of the participants in the TFT condition. The participants whose data are shown in the left graph "escaped" the trap of mutual defections versus TFT. The criteria for choosing these participants were: (a) cooperation on fewer than half of the trials in the first half of the session, and (b) cooperation on more than half of the trials in the second half. The participants whose data are shown in the right graph were "trapped" in the pattern of mutual defection. The criteria for choosing these participants were: (a) cooperation on fewer trials in the second half of the session than in the first half, and (b) cooperation on fewer than half of the trials in the second half.

Data from the RAN group were examined to determine the probability of cooperation as a function of the feedback on the previous trial. The total number of cooperative choices following trials with a particular feedback (when 2 DP's cooperated/2 defected, for instance) was divided by the total number of choices following those trials. The top row of Table 1 shows the overall probability of cooperation given the five feedback possibilities. A cooperation was least likely to follow a trial in which all 4 DPs defected, and this probability climbed as the number of DPs cooperating on the previous trial increased up to the point when 3 DPs cooperated/1 defected, and then dropped when all 4 DP's cooperated. The second and third rows of Table 1 show data from the same players, but divide the session into first and second halves: The same pattern of cooperation was evident within both halves of the session.

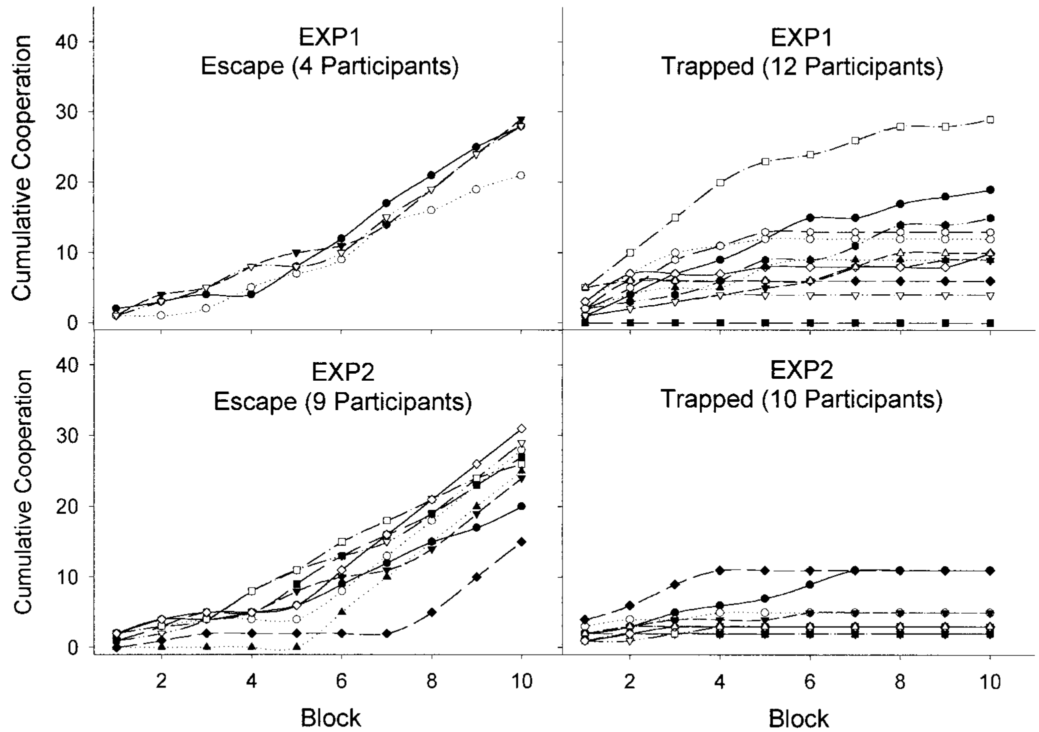

Fig. 6. Cumulative cooperation counts of those classified as those who escape (left) and are trapped (right) in Experiments 1 (top) and 2 (bottom).

Discussion

The overall level of cooperation in the RAN group, by the end of the experiment, was not significantly different from that of the preliminary experiment in which participants played each other. As in the preliminary experiment, the number of other players cooperating in the RAN group could shift widely from trial to trial (whereas in the TFT condition the number of DPs cooperating could not increase or decrease by more than 1 from trial to trial).

As expected, there was no significant difference in mean cooperation rate between players in the TFT and RAN groups early in the session. However, the moderate difference between means in the first block combined with the similarity in number of players cooperating/defecting on the first trial suggests that the two reinforcement contingencies started influencing behavior almost immediately. This may simply be due to a difference between TFT and RAN on the first trial: With TFT, 4 DPs always cooperated on the first trial; with RAN, there was only a .2 probability that all DPs would cooperate. Given that most players in the RAN group faced at least some defection by DPs on the first trial, the higher cooperation rate by the TFT group could be due to a tendency by players to reciprocate cooperation (at least early in the session).

Overall, the mean cooperation rate was

Table 1

Probabilities of cooperation in the random condition (RAN) given the distribution of dummy player (DP) choice in the previous trial.

	Number of DP's defecting/ Number of DP's cooperating				
	4/0	3/1	2/2	1/3	0/4
Experiment 1					
50 trials	.22	.22	.26	.46	.26
First 25 trials	.23	.25	.23	.35	.32
Second 25 trials	.19	.20	.28	.57	.20
Experiment 2					
Phase I	.23	.25	.24	.37	.28
Phase II	.08	.12	.22	.12	.16

higher for players in the TFT group than in the RAN group. This was expected, as was the decrease observed across the session for the RAN group. Mean cooperation versus TFT was moderate early in the session and remained so throughout. Although the expected increase in cooperation by the TFT group was not obtained within the session, mean cooperation rate is consistent with previous data. Experiment 8 of Rachlin et al. (2001) was a two-player IPDG similar to this experiment in that players were given the cover story of a social dilemma although they were essentially in a self-control situation. The mean cooperation rate in their version of TFT was higher than that obtained in the present experiment (~70% vs. 46%). Given the greater number of players in the current INPDG (hence more diffuse DP reciprocation), the moderate cooperation rate is not surprising.

The frequency distributions of Figure 5 suggest that the RAN condition decreased cooperation in nearly all players. The number of exclusive defectors increased from 6 in Block 1 to 20 by Block 10, and only 4 participants cooperated more than two times in the last block. Furthermore, 21 participants in the RAN condition (not shown) met the operational definition for "trapped," whereas only 2 met the criteria for "escape." Although the distribution of the TFT group became bimodal in Block 10, the number of exclusive cooperators did not increase from Block 1 to Block 10. In fact, 5 of the 7 points that represent exclusive cooperation in both Blocks 1 and 10 represent individual players who cooperated exclusively for the entire session. The existence of the 4 escapers (Figure 6) indicates, however, that some players did increase their cooperation rates enough to increase overall amount earned.

Table 1 shows that, for the RAN group, the behavior of DPs strongly influenced cooperation even though the best strategy versus RAN was to defect on all trials. Up to a point, this influence was direct; the more DPs cooperated, the more the real players cooperated. But this was not simply probability matching because, when all DPs cooperated, players versus RAN tended to defect on the next trial, and this tendency was accentuated as the session progressed. A player could follow a loose TFT rule: Cooperate after a trial in which many DPs cooperated (and thus reward "others" for cooperating); defect after many had defected (and thus punish "others" for defecting). This is consistent with Erev and Roth's (2001) finding that a "forgiving" TFT strategy models human IPDG performance. This is also consistent with behavior observed by Fox and Guyer (1977), who found more cooperation in a three-player INPDG when the probability of cooperation by other players was relatively high (64%) than low (36%).

If a player in the current experiment exposed to TFT used a reciprocal strategy but defected on a trial because of error, confusion, or momentary impulsiveness, the player's defection would have been immediately reciprocated by 1 DP. The player could then reciprocate defection by 1 DP and deterioration to complete defection would result (Kollock, 1993; Macy, 1996). Although this deterioration to mutual defection would take a number of trials in the current experiment, players sensitive to being exploited by even a single DP would come to defect exclusively. Additionally, individual defections would be reinforced (as always with PDGs), whereas full punishment (all or most DPs defecting) would be delayed for a number of trials. Following a string of defections, the cost of cooperating may have been too high to counterbalance the immediate reinforcement of defection. With DPs and the participant all applying TFT, a participant who varied even slightly from exclusive cooperation could become hopelessly trapped in exclusive defection—unless she or he, at least temporarily, cooperated in the absence of DP cooperations. A player in this situation would have to cooperate for at least two trials before seeing any improvement in overall earnings relative to those obtained by repeated defection.

The trap at the point of repeated defection would be even tighter if reward magnitude were translated to reward value. Assuming diminishing marginal value (value as a log function of magnitude, for example), the distance between the solid diagonal lines of Figure 2 would be greater at the lower right than at the upper left, implying that the immediate punishment for cooperation would be greater after a string of defections than after a string of cooperations—putting together such a string of cooperations, hence rising to the left along the broken line of Figure 2, would be more and more difficult the further to the right in Figure 2 that the string began.

A further implication of diminishing marginal value is that release from the trap by arbitrarily moving an INPDG player to the leftmost point of Figure 2 (forgiving past defections) would reduce the immediate punishment of cooperation thereby making it easier for that player to begin a string of cooperations and obtain the reward for so doing. Experiment 2 was designed to test this implication.

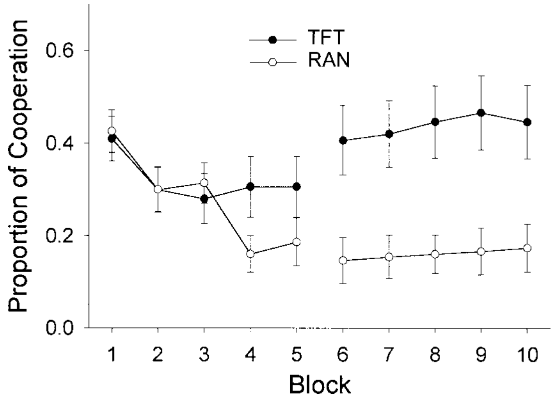

Fig. 7. The mean cooperation rate in 5 five-trial blocks in the two phases versus TFT (filled circles) and RAN (open circles) of Experiment 2. Error bars indicate standard error of the mean.

EXPERIMENT 2

Prior defection against TFT causes current cooperation to be punished initially; only a succession of cooperations is reinforced relative to exclusive defection. In a new game, however, the consequences of prior defections are removed. A player who had defected early in the session in Experiment 1 would have been choosing between the alternatives at the right of Figure 2. In a new game against TFT, those alternatives would suddenly shift back to the initial alternatives at the left. Cannon and McSweeney (1998) found that a midsession pause in the reinforcement schedule created a "restarting" of the within-session pattern of responding when the session continued. It is possible that a player who would have continued defecting without resetting would cooperate when starting anew.

METHOD

Participants

Sixty participants (26 male) were recruited from the Psychology Department participant pool at the University at Stony Brook.

Procedure

One group of 30 participants was exposed to the TFT condition in two separate 25 trial phases (I and II). The second group was exposed to the RAN condition in both phases. Participants were informed that the experiment would entail two phases. Directions were the same as those of Experiment 1 in all other respects. At the end of Phase I, all participants were given the distracter questionnaire while the experimenter reset the program for Phase II. Once the program had been reset and the distracter task was complete, participants were directed to "begin Phase II." All DPs cooperated on the first trial of Phase II in the TFT condition, whereas the number of DPs cooperating on that trial was randomly determined in the RAN condition. As in Experiment 1, feedback was presented 5 s after all players had submitted their choices and was kept on the screen for 15 s. At the conclusion of Phase II, participants filled out the postexperiment questionnaire. Participants were debriefed and asked not to discuss the nature of this experiment with others. They were then paid the amount they earned in the experiment and dismissed.

RESULTS

Trials in both conditions were divided into 5 five-trial blocks and mean proportion of cooperation for each block was calculated for each phase and condition (Figure 7). Cooperation rate was similar between TFT and RAN in Phase I: Cooperation decreased within the first five blocks, though more so in the RAN than in the TFT condition. The difference between strategy conditions was more apparent in Phase II, where more cooperation was observed in the TFT than RAN condition. Cooperation rate appeared to increase in Phase II in the TFT condition (relative to Phase I), whereas it decreased slightly in the RAN condition. A $2 \times 2 \times 5$ ANOVA was conducted to examine these effects (condition, phase, and block).

There was a main effect of block, $F(4, 232) = 3.21$, $p < .05$, and no main effect of phase,

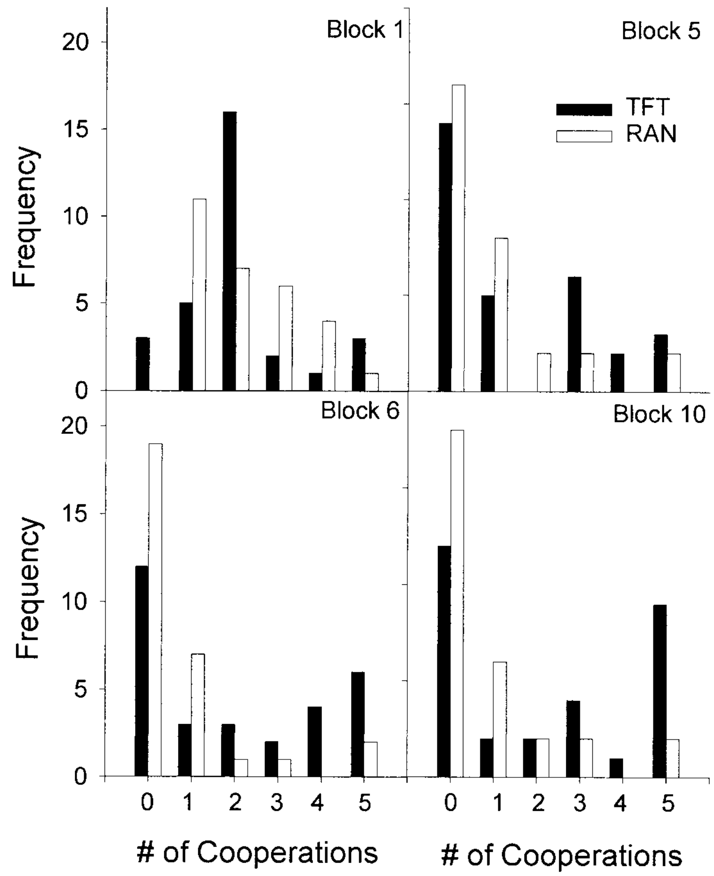

Fig. 8. The number of participants as a function of number of cooperations in Blocks 1, 5, 6, and 10 versus TFT (filled bars) and RAN (open bars) of Experiment 2.

F (1, 58) = 0.00, $p > .05$. More important, there was a significant difference in overall mean proportion of cooperation between the TFT ($M = 0.38$) and RAN ($M = 0.22$) groups, F (1, 58) = 6.52, $p < .05$. The condition x phase interaction was significant, F (1, 58) = 11.82, $p < .05$. Planned comparisons showed that the difference in Phase I between TFT ($M = 0.32$) and RAN ($M = 0.28$) was not significant, F (1, 58) = .54, $p > .05$. In Phase II, more cooperation was observed in TFT ($M = 0.44$) than RAN ($M = 0.16$), F (1, 58) = 11.55, $p < .05$. The mean proportion of cooperation in the RAN group decreased between Phase I and II, F (1, 58) = 5.98, $p < .05$; the mean proportion of cooperation in the TFT group increased between phases, F (1, 58) = 5.84, $p < .05$.

Figure 8 displays frequency distributions of the number of participants cooperating in Blocks 1, 5, 6, and 10. As in Figure 5, the frequency distributions of the first block of both TFT and RAN groups are somewhat bell-shaped, and 27 of 30 and 30 of 30 players, respectively, cooperated at least once in the first block. As the session progressed, the shape of the distribution in TFT became bimodal, with the major mode at 0 (12 observations) and a minor mode at 5 (nine observations) by Block 10. The majority of participants were cooperating or defecting exclusively by the end of the session in the TFT condition. The shape of the distribution in RAN quickly became positively skewed (mode at 0) as the session progressed. The majority of these participants defected for most of the session.

The bottom graphs of Figure 6 show cumulative cooperations of TFT participants who escaped (left graphs) or were trapped (right graphs). The selection criteria were the same as those used in Experiment 1. Nine participants escaped the trap of repeated defection, whereas 10 were trapped in repeated defections. For some of those who escaped, the increase in cooperation in Phase II was dramatic, including three instances in which defection on most trials in Phase I was followed by exclusive cooperation in Phase II. One of these participants defected exclusively in Phase I and cooperated exclusively in Phase II. For most of the trapped participants, overall cooperation rate was low with very few cooperations after the first few blocks.

The bottom rows of Table 1 show probabilities of cooperation following each distribution of DP choices on the previous trial for the RAN group in Phases I and II. Overall, data were similar to those of Experiment 1; probability of cooperation showed some increase as the number of DPs cooperating on the previous trial increased. As in Experiment 1, the highest probability of cooperation did not follow trials in which all 4 DPs cooperated; rather the highest probabilities of cooperation followed trials in which 3 DPs cooperated in Phase I or 2 DPs cooperated in Phase II. In the RAN condition, the probability of cooperation generally was higher earlier in the session (Phase I) than later (Phase II).

Discussion

As in Experiment 1, a higher mean cooperation rate was observed in the TFT condition than the RAN condition. Observations from the RAN condition were nearly identical to those of Experiment 1, where cooperation rate decreased immediately and remained low after reaching about 20% cooperation. Observations from the TFT condition were somewhat different: Following an initial decrease in cooperation, a significant increase was observed in Phase II. Frequency distributions indicate that mean cooperation rate in TFT was maintained because, whereas some players defected exclusively, many also cooperated exclusively (mean skewness = 0.63). On the other hand, most players in RAN defected exclusively (mean skewness = 1.60).

In Phase II, there was a significant difference between mean cooperation rate for the TFT and RAN groups. Previous exposure to RAN decreased mean cooperation rate in later exposure to RAN, particularly early in Phase II. In addition, previous exposure to TFT increased mean cooperation rate in later exposure to TFT. This is consistent with the bimodal frequency distribution of Experiment 1: if the player applies some reciprocal strategy in the presence of TFT, one accidental defection can quickly lead to deterioration or to exclusive defection. This could have occurred for some players in Phase I. Given a second chance in the second phase of the experiment, those players may have been less likely than before to emit the first nonreciprocal response. The increase in number of exclusive cooperators in Phase II (compared to Phase I) lends support to this idea. Furthermore, the number of participants who are classified as having escaped, as well as the plots of their cumulative cooperation (bottom-left graph of Figure 6), points to specific instances in which this pattern of responses occurred.

Figure 6 compares individual performance in Experiments 1 and 2. Though some participants in both experiments escaped the trap of repeated defections, the escapes of participants in Experiment 2 were more numerous and dramatic. A number of participants in Experiment 2 who were not cooperating at all at the end of Phase I exclusively cooperated for the duration of Phase II. Behavior of this type was not seen in Experiment 1. The differences between the outcomes of the two experiments appear in the second half of the session. This was expected because up to Trial 25 of the experimental session, the contingencies in the two experiments were identical.

EXPERIMENT 3

Two elements differentiated the TFT conditions of Experiments 1 and 2. First, there was a clear temporal distinction between the two halves of the session in Experiment 2 that did not occur in Experiment 1. Second, a resetting occurred in Experiment 2, which made the distribution of DP choices at the start of Phase II completely independent of player choices in Phase I. These two elements were evaluated separately to determine which was crucial for the increase in cooperation observed in the second phase of Experiment 2.

Method

Participants

Sixty participants (20 male) were recruited from the Psychology Department participant pool at the University at Stony Brook.

Procedure

All 60 participants were exposed to the TFT condition for a total of 50 trials. Half of the participants (RESET condition) experienced all 50 trials without a pause, similar to Experiment 1. Following the 25th trial, however, the TFT strategy was reset (without a temporal marker or players' knowledge) such that all DPs cooperated on the 26th trial. The strategy then followed the normal TFT strategy for the remainder of the session. The remaining half of the participants (STOP condition) experienced a pause in the game halfway through the session, similar to Experiment 2. The TFT strategy was not reset, however, and the distribution of DP choices on the 26th trial followed what was determined from the combined effects of the player's choice and the distribution of DP choices on the 25th trial. As in the previous experiments, feedback was presented 5 s after all players had submitted their choices and was kept on the screen for 15 s. Other than the experimental condition, contingencies in this experiment were identical to those of Experiment 2.

Results

Trials in both conditions were divided into 10 five-trial blocks; analyses were conducted on mean cooperation rate for each block for each condition (Figure 9). A 2 × 2 × 5 mixed ANOVA was conducted to examine the effect of condition, phase, and block on mean cooperation rate. There was a main effect of phase, with more cooperation observed in Phase II ($M = 0.36$) than Phase I ($M = 0.29$), $F\,(1, 58) = 4.94$, $p < .05$. No differences were observed between the STOP ($M = 0.35$) and RESET conditions ($M = 0.31$), $F\,(1, 58) = 0.29$, $p > .05$, nor between blocks, $F\,(4, 232) = 1.21$, $p > .05$. None of the interactions were significant.

Planned comparisons were conducted to examine further the within-group differences

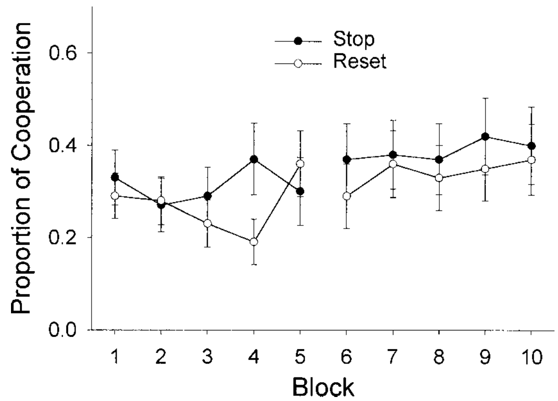

Fig. 9. The mean cooperation rate in 5 five-trial blocks in the two phases versus TFT in the STOP (filled circles) and RESET (open circles) conditions of Experiment 3. Error bars indicate standard error of the mean.

between phases. In the STOP condition, there was a nonsignificant difference in mean cooperation rate between Phases I ($M = 0.31$) and II ($M = 0.39$), $F\,(1, 58) = 2.70$, $p > .05$. The difference between Phases I ($M = 0.27$) and II ($M = 0.34$), $F\,(1, 58) = 2.24$, $p > .05$, in the RESET condition also was nonsignificant. Frequency distributions showed the pattern observed in the previous experiments: somewhat normal distributions became increasingly bimodal as the session progressed, with the major mode at 0 and the minor mode at 5.

Figure 10 shows cumulative cooperations of participants considered to have escaped or to have been trapped in the two conditions. The numbers of participants who escaped the pattern of repeated defections is comparable for the two groups; these numbers are greater than those of the TFT group of Experiment 1 and less than those of the TFT group of Experiment 2.

Discussion

There was an overall increase in cooperation between Phases I and II, but the increases observed in each of the two conditions separately were not statistically significant. This is dissimilar to the result of Experiment 2, in which the TFT condition resulted in a significant increase in mean cooperation rate in the second phase of the experiment.

The pattern of results obtained in the STOP and RESET conditions are nearly identical to each other, and somewhat similar to

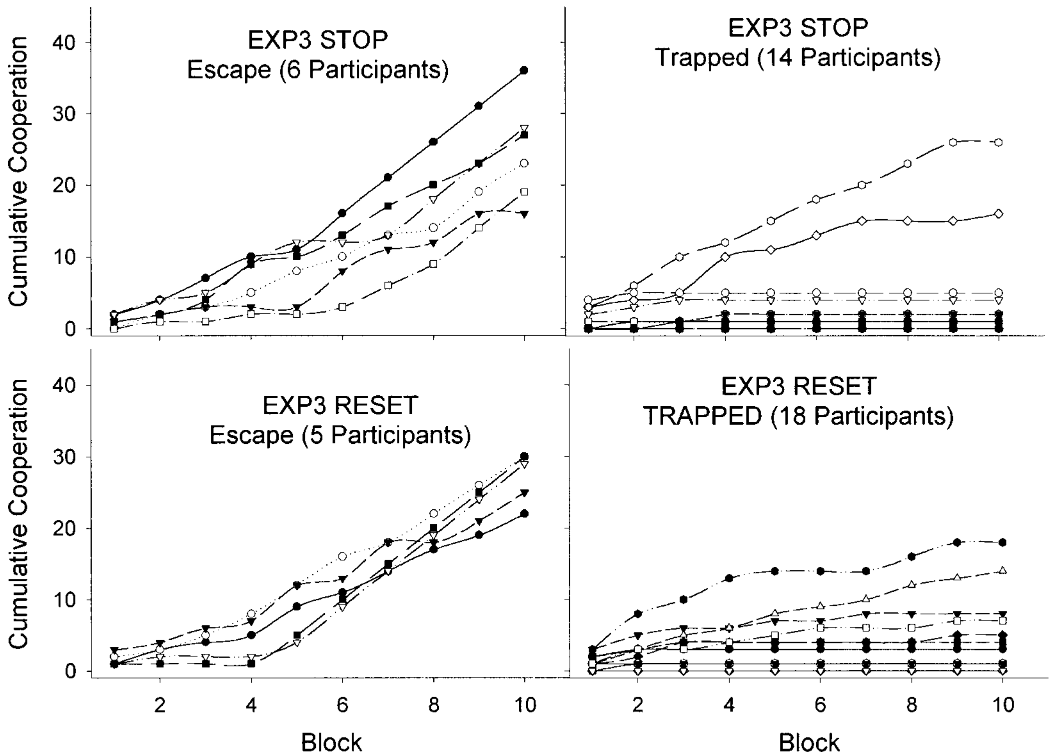

Fig. 10. Cumulative cooperation counts of those classified as those who escape (left) and are trapped (right) in the STOP (top) and RESET (bottom) conditions of Experiment 3.

the TFT condition of Experiment 2. The major difference between the TFT condition of Experiment 2, which includes both a stop and reset, and the STOP and RESET conditions of Experiment 3, is the degree to which

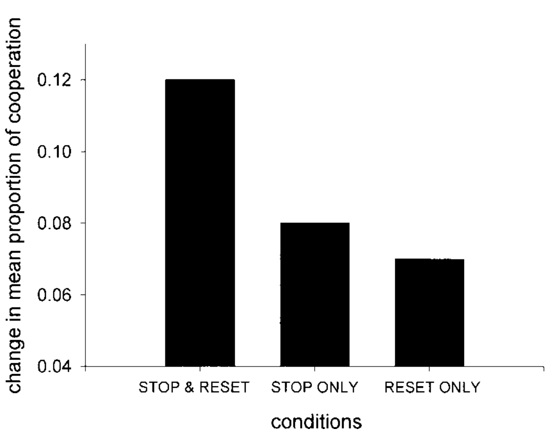

Fig. 11. The increase in mean cooperation rate per block between phases versus TFT in three conditions differing in the manner in which the phases are divided (from Experiments 2 and 3).

cooperation increased in Phase II relative to Phase I. Figure 11 displays the change in mean cooperation rate from Phases I to II in the three conditions. The TFT condition of Experiment 2 is distinguished by a greater increase in cooperation between phases than either of the Experiment 3 conditions, though this observed difference is not statistically significant, F (2, 87) $= 0.25$, $p > .05$. An examination of the bottom of Figure 6 and Figure 10 indicates that the overall increase in cooperation, though nonsignificant, is due to the distribution of participants who mostly/exclusively cooperate and mostly/exclusively defect. The stop and resetting of the TFT strategy of Experiment 2 appears to allow more participants to escape the trap of defections than either the STOP or RESET conditions of Experiment 3.

CONCLUSIONS

The manner in which TFT was applied in the current social cooperation study models an established choice procedure in studies of self-control (Herrnstein et al., 1993; Kudadjie-Gyamfi & Rachlin, 1996). In the present social cooperation studies, as in self-control studies, a conflict was imposed between maximization of local and global rewards. The main difference between the two types of studies is in the instructions to the participants. On the one hand, in the present experiments participants were told that they were playing against other real players but were actually playing against dummy players. The players, therefore, did not know the rules by which their play influenced the choices of dummy players. However, they were told the rules by which "other players' choices" influenced the alternatives they faced. On the other hand, participants in typical self-control studies are correctly told that they are interacting with a computer and that reward depends on their own behavior, but they are not explicitly informed of the contingencies between their behavior and reward. In neither type of study are participants typically able to state those contingencies after exposure to them over the limited number of experimental trials. Nevertheless, in both types of studies, behavior is at least partially responsive to them—self-control is higher, hence overall reinforcement rate is higher, for participants exposed to such contingencies than those exposed to random variation of reward, albeit few participants with either set of instructions maximize global reinforcement.

Experiment 2 shows that, with the prisoner's dilemma, when the trap of mutual defection is opened, and players are given a new chance to cooperate, cooperation does increase. Although the likely explanation for this effect is a new opportunity to cooperate versus a reciprocal strategy (as opposed to a random strategy), the two experimental conditions also differed in one additional way. In each phase of the TFT condition, all DPs cooperated on the first trial. This was not the case in the first trial of each phase in the RAN condition; the number of DPs cooperating was randomly determined and, on average, only 1 of 5 players in the RAN condition had 4 DPs cooperating on the first trial. However, it is unlikely that this first-trial difference was the critical difference between conditions; in previous research, PDG participants did not increase cooperation even when their opponents cooperated on *all* trials (Silverstein, Cross, Brown, & Rachlin, 1998).

The STOP condition of Experiment 3 further discounts the possibility that the number of DPs cooperating on the first trial of each phase was the critical difference between conditions. In the STOP condition, unlike the TFT condition of Experiment 2 and the RESET condition of Experiment 3, the number of DPs cooperating on the first trial of Phase II was not 4 for all players. Because that number depended on each participant's behavior in the previous trials, the number of DPs cooperating in the first trial of Phase II was frequently low. Nonetheless, without an all-cooperate condition (or a RAN condition in which 4 DPs always cooperated on the first trial), first-trial difference cannot be definitively ruled out as an explanation for heightened cooperation versus TFT.

Experiment 3 indicates that the increase in cooperation in the NPDG versus TFT depended not only on a resetting of contingencies but also on a clear signal (here a pause in the game) that they had been reset. It remains to be determined whether self-control in situations of complex ambivalence may be similarly enhanced.

Although there is some evidence from the present studies that participants themselves played a modified form of tit for tat (they reciprocated other players' cooperation, but with a low probability), a preliminary experiment with multiple real players found a very low level of mutual cooperation (about equal to that against randomly playing opponents). This may be due to the low probability of reciprocation of cooperation by real players. Reciprocation by the tit-for-tat playing DPs in the present experiments (although graded rather than all-or-none) was certain and did engender cooperation.

These experiments thus imply that cooperation in real-life social systems with multiple participants will be enhanced when reciprocation is certain, when multiple defections by any given group member are periodically forgiven by the other members, and when that forgiveness is clearly signaled.

REFERENCES

Ainslie, G. (1992). *Picoeconomics.* New York: Cambridge University Press.

Axelrod, R. (1984). *The evolution of cooperation.* New York: Basic Books.

Cannon, C. B., & McSweeney, F. K. (1998). The effects off stopping and restarting a session on within-session patterns of responding. *Behavioural Processes, 43,* 153–162.

Erev, I., & Roth, A. E. (2001). Simple reinforcement learning models and reciprocation in the prisoner's dilemma game. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 215–231). Cambridge, MA: MIT Press.

Fox, J., & Guyer, M. (1977). Group size and others' strategy in an n-person game. *Journal of Conflict Resolution, 21,* 323–338.

Harris, A. C., & Madden, G. J. (2002) Delay discounting and performance on the prisoner's dilemma game. *Psychological Record, 52,* 429–440.

Herrnstein, R. J., Loewenstein, G. F., Prelec, D., & Vaughan, W. (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioral Decision Making, 6,* 149–185.

Heyman, G. M. (2003). Consumption dependent changes in reward value: A framework for understanding addiction. In R. E. Vuchinich & N. Heather (Eds.), *Choice, behavioural economics and addiction* (pp. 95–121). Oxford, England: Pergamon.

Kollock, P. (1993). ''An eye for an eye leaves everyone blind'': Cooperation and accounting systems. *American Sociological Review, 58,* 768–786.

Komorita, S. S. (1976). A model of the N-person dilemma-type game. *Journal of Experimental Social Psychology, 12,* 357–373.

Komorita, S. S., Parks, C. D., & Hulbert, L. G. (1992). Reciprocity and the induction of cooperation in social dilemmas. *Journal of Personality and Social Psychology, 62,* 607–617.

Kudadjie-Gyamfi, E., & Rachlin, H. (1996). Temporal patterning in choice among delayed outcomes. *Organizational Behavior and Human Decision Processes, 65,* 61–67.

Kudadjie-Gyamfi, E., & Rachlin, H. (2002). Rule-governed versus contingency-governed behavior in a self-control task: Effects of changes in contingencies. *Behavioural Processes, 57,* 29–35.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (pp. 111–194). Princeton, NJ: Princeton University Press.

Macy, M. (1996). Neural selection and social learning in the prisoner's dilemma: Co-adaptation with genetic algorithms and neural networks. In W. B. G. Liebrand & D. M. Messick (Eds.), *Frontiers in social dilemma research* (pp. 235–266). Berlin: Springer-Verlag.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 5. The effect of delay and of intervening events on reinforcement value* (pp. 55–73). Hillsdale, NJ: Erlbaum.

Rachlin, H. (2000). *The science of self-control.* Cambridge, MA: Harvard University Press.

Rachlin, H., Brown, J., & Baker, F. (2001). Reinforcement and punishment in the prisoner's dilemma game. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory, 40* (pp. 327–364). San Diego, CA: Academic Press.

Silverstein, A., Cross, D., Brown, J., & Rachlin, H. (1998). Prior experience and patterning in a prisoner's dilemma game. *Journal of Behavioural Decision Making, 11,* 123–138.

Received February 4, 2004
Final acceptance July 1, 2004