ELSEVIER

Behavioural Processes 57 (2002) 211–226

BEHAVIOURAL PROCESSES

www.elsevier.com/locate/behavproc

Teaching and learning in a probabilistic prisoner's dilemma

Forest Baker, Howard Rachlin *

Department of Psychology, State University of New York at Stony Brook, Stony Brook, NY 11794, USA

Accepted 16 November 2001

Abstract

The prisoner's dilemma is much studied in social psychology and decision-making because it models many real-world conflicts. In everyday terms, the choice to 'cooperate' (maximize reward for the group) or 'defect' (maximize reward for the individual) is often attributed to altruistic or selfish motives. Alternatively, behavior during a dilemma may be understood as a function of reinforcement and punishment. Human participants played a prisoner's-dilemma-type game (for points exchangeable for money) with a computer that employed either a teaching strategy (a probabilistic version of tit-for-tat), in which the computer reinforced or punished participants' cooperation or defection, or a learning strategy (a probabilistic version of Pavlov), in which the computer's responses were reinforced and punished by participants' cooperation and defection. Participants learned to cooperate against both computer strategies. However, in a second experiment which varied the context of the game, they learned to cooperate only against one or other strategy; participants did not learn to cooperate against tit-for-tat when they believed that they were playing against another person; participants did not learn to cooperate against Pavlov when the computer's cooperation probability was signaled by a spinner. The results are consistent with the notion that people are biased not only to cooperate or defect on individual social choices, but also to employ one or other strategy of interaction in a pattern across social choices. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Prisoner's dilemma; Cooperation; Strategies; Tit-for-tat; Pavlov; Instructions; Context; Humans

1. Introduction

Fig. 1(a) illustrates a two-person prisoner's dilemma game. Each of two players (A and B) chooses between two alternatives (C and D). If both players 'cooperate' (choose C), each obtains a reward of 5 U; if both 'defect' (choose D), each obtains a reward of 2 U. However, if one player cooperates while the other defects, the cooperator obtains only 1 U, while the defector obtains 6 U.

The dilemma posed by the game is a conflict between the choice (defect) that maximizes each individual player's reward and the choice (cooperate) that maximizes reward for the group as a whole. Consider the game of Fig. 1(a) from the viewpoint of Player A: if Player B cooperates, Player A's reward is constrained to the upper two boxes; 5 U for cooperating versus 6 U for defecting; if Player B defects, Player A's reward is confined to the lower two boxes; 1 U for cooperating versus 2 U for defecting. Regardless of whether Player B cooperates or defects, Player A's reward is 1 U higher for defecting than for cooperating. Player A therefore maximizes reward by

* Corresponding author. Tel.: +1-631-632-7807; fax: +1-631-632-7876.

E-mail address: howard.rachlin@sunysb.edu (H. Rachlin).

0376-6357/02/$ - see front matter
PII: S0376-6357(02)00015-3

defecting. Since the game is perfectly symmetric, Player B also maximizes reward by defecting. Now consider the game from the viewpoint of the group as a whole: if both players cooperate, the total reward is 10 U (5 + 5); if one player cooperates while the other defects, the total reward is 7 U (1 + 6); if both players defect, the total reward is 4 U (2 + 2). Thus, from the viewpoint of the group as a whole, reward is maximized by cooperation.

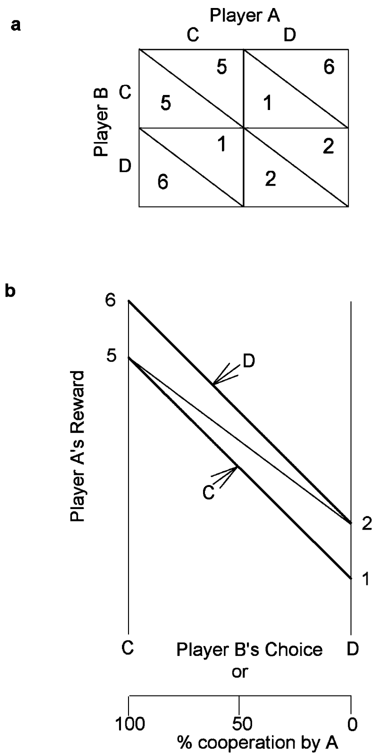

Fig. 1. (a) Contingency matrix of a typical two-player prisoner's dilemma game. The numbers in the matrix represent amounts of reward for two players (A and B) contingent on each player's choice between two alternatives: cooperate (C) or defect (D). If both players cooperate, each gets five points; if both defect, each gets two points. If one player cooperates while the other defects, the cooperator gets one point while the defector gets six points. (b) Repeated prisoner's dilemma game (with the matrix of (a)) from the viewpoint of Player A. Player B's choice (C or D) is represented by the two vertical lines. If Player B cooperates (left vertical line), Player A earns six points by defecting and five by cooperating; if Player B defects (right vertical line), Player A earns two points by defecting and one by cooperating. The thin diagonal line indicates Player A's average earnings against tit-for-tat as a function of percent cooperation on repeated trials versus tit-for-tat.

The prisoner's dilemma is much studied in social psychology and decision-making because it models many real-world conflicts, ranging from littering (defection = littering; cooperation = not littering) to international arms races (defection = arming; cooperation = disarming) (for reviews see Dawes, 1980; Messick and Brewer, 1983; Van Lange et al., 1992; Komorita, 1994; Komorita and Parks, 1995; Kollock, 1998). In laboratory studies, humans playing prisoner's dilemma games typically defect (Rapoport and Chammah, 1965). Such defection has usually been taken as evidence of a bias towards 'selfishness' over 'altruism' (Caporael et al., 1989). This way of expressing the dilemma, however, is highly artificial. The maximization of individual reward consequent on defection supposes that the game is played only once and outside of any social context. In actual real-world instances, however, games are played repeatedly and in the context of prior and subsequent social interaction. When games are played repeatedly, various strategies, extending over multiple instances, can engender cooperation (Rapoport, 1967; Komorita et al., 1991; Silverstein et al., 1998; Rachlin et al., 2000).

Another way of looking at the prisoner's dilemma focuses on contingencies of reinforcement and punishment over repeated trials (Rachlin et al., 2000). Fig. 1(b) (thick diagonal lines) shows the game of Fig. 1(a) from the viewpoint of Player A. If, on a given trial, Player B cooperates, Player A next may choose between 5 and 6 U; if player B defects, Player A next may choose between only 1 and 2 U. Therefore, Player A's available alternatives will be much greater in value if Player B cooperates than if Player B defects. As Fig. 1(b) shows, Player A's own choice to defect rather than cooperate gains an average of 1 U (the vertical distance between the two heavy diagonal lines), whereas Player B's cooperation rather than defection gains Player A an average of 4 U (5 or 6 versus 1 or 2 U). Player B's cooperation may therefore be a very strong reinforcer of Player A's choices in repeated trials. Because the game is symmetric, Player A's cooperation also reinforces Player B's choices. Correspondingly, defection by either player is a strong punisher of the other player's prior choices. The choices of either player thus reinforce or punish the choices of the other; in that sense, each player teaches the other. Simultaneously, the choices of either player are reinforced by or punished by the choices of the other; in that sense, each player learns from the other. These two roles, teacher and learner, imply two strategies found to strongly influence choice in the prisoner's dilemma game: a strategy that teaches the other player ('tit-for-tat'; Axelrod, 1984) and a strategy that learns from the other player ('Pavlov'; Fudenberg and Maskin, 1990; Nowak and Sigmund, 1993). The purpose of the present experiments was to compare the effectiveness of these strategies in producing cooperation in repeated prisoner's dilemma games and to test the generality of these effects in different contexts.

Numerous studies have used computer-based evolutionary game theoretic models to test which strategy (tit-for-tat or Pavlov) earns a higher rate of reinforcement; that is, induces more cooperation from other simulated strategies (see Kraines and Kraines, 1989; Fudenberg and Maskin, 1990; Nowak and Sigmund, 1993; Kraines and Kraines, 1993, 1995; Macy, 1995; Messick and Liebrand, 1995). To our knowledge, the present study is the first to test the level of cooperation that the Pavlov strategy induces from real human participants.

1.1. Strategies in repeated prisoner's dilemma games

The strategy of tit-for-tat reinforces the other player's cooperation on trial n by cooperating on trial $n+1$ and punishes the other player's defection on trial n by defecting on trial $n+1$. This strategy, consistently applied, has been found to produce mutual cooperation by human participants over a series of trials (Komorita, 1994). If Player B plays tit-for-tat consistently, the label 'Player B's Choice' in Fig. 1(b) may be replaced by 'Player A's Previous Choice.' Thus, against tit-for-tat, Player A is essentially in a self-control situation (Ainslie, 1992; Rachlin, 2000), choosing between a small, immediate reward—a 1-U reward difference on the present trial (obtained by defecting on the present trial) and a larger but delayed reward—a 4-U reward difference on the next trial (obtained by cooperating on the present trial). The long-term benefits of cooperating are shown by the thin solid diagonal line in Fig. 1(b.) The line indicates average reward as a function of percent cooperation over repeated trials against tit-for-tat (scale at the bottom). The highest average reward (5 U per trial) is attained with 100% cooperation; reward decreases proportionally as defections increase to the lowest average reward (2 U per trial) with 100% defection.

Because trials in laboratory studies with human subjects are usually only a few seconds apart, the delay to the larger reward may have little effect in reducing Player A's cooperation when Player B employs tit-for-tat. On the other hand, in studies with non-humans playing against tit-for-tat, where delays of a few seconds may significantly decrease reward value, cooperation has been difficult to achieve (Green et al., 1995). However, by delaying the reward for defecting to equal that for cooperating, Stephens (2000) has found that bluejays (playing against 'stooge' opponents trained to follow a pattern that effectively simulated tit-for-tat) cooperated consistently.

Another prisoner's dilemma strategy, one that puts the strategy's user (Player B) in the position of a learner rather than a teacher, is called Pavlov. (Because the learning process of this strategy is instrumental, it should have been called 'Skinner' or 'Thorndike'.) A player using Pavlov cooperates on trial $n+1$ if both players either cooperated or both players defected on trial n, but defects on trial $n+1$ if one player cooperated and the other defected on trial n. In effect, if Player B uses Pavlov, Player A's cooperation will reinforce Player B's prior choice (will cause Player B to repeat the prior choice) while Player A's defection will punish Player B's prior choice (will cause Player B to switch from the prior choice). Table 1 illustrates the Pavlov strategy and compares it with tit-for-tat.

Tit-for-tat is a simpler strategy than Pavlov. Using tit-for-tat, Player B simply mimics Player A's prior choice; in Table 1, the third column matches the second. Using Pavlov, Player B's choice on trial $n+1$ depends on both Player A's

Table 1
Comparison of the strategies of tit-for-tat and Pavlov

	Trial n		Computer's choice on trial $n+1$	
	Computer's choice	Subject's choice	When using tit-for-tat	When using Pavlov
1	C	C	C	C
2	D	C	C	D
3	C	D	D	D
4	D	D	D	C

C, cooperate; D, defect.

prior choice and B's own prior choice; if A cooperated (rows 1 and 2), B's prior choice is repeated; if A defected (rows 3 and 4), B's prior choice is changed. Note that tit-for-tat and Pavlov dictate the same choice in two of the four cases: when both players cooperate on trial n (row 1) and when A defects while B cooperates on trial n (row 3). However, only consistent cooperation by both players (row 1) is stable with both strategies. If A consistently defects over a series of trials, B, using tit-for-tat, would defect in turn (rows 3 and 4); but, in the face of consistent defection by A over a series of trials, B, using Pavlov, would switch back and forth between cooperation and defection (between rows 3 and 4). In computer tournaments Pavlov has outperformed (earned greater reward than) all other strategies, including tit-for-tat (Kraines and Kraines, 1989, 1993, 1995; Nowak and Sigmund, 1993; Messick and Liebrand, 1995).

The differing choices indicated in the cells of Table 1 indicate that, unless Player A is cooperating on all trials, Player B cannot simultaneously employ both tit-for-tat and Pavlov. Yet, in a two-player game, each player is both a learner and a teacher; that is, each player's choices both reinforce and punish and are reinforced and punished by the choices of the other player. The incompatibility of teaching and learning may account for failures to reach consistent cooperation in laboratory studies of repeated prisoner's dilemma games (Silverstein et al., 1998).

The present experiments differed from typical prisoner's dilemma studies in two ways. First, the computer and the participant chose alternately rather than simultaneously. With the usual simultaneous procedure, the effect of a player's choice on the other player's choice and the effect of a player's choice on reward are confounded. If the players choose alternately, however, these two effects are more discriminable. Moreover, in real-life prisoner's-dilemma-type situations (arms races for example), choices are often made alternately (as 'tit-for-tat' implies).

A second difference from typical prisoner's dilemma studies is that in the present study the computer's cooperation or defection varied probabilistically. The computer began the experiment by cooperating (as is standard practice) but thereafter, instead of cooperating or defecting, as determined from Table 1, the computer (starting with a base cooperation probability of 50%) increased or decreased its cooperation probability by 25% (defection probability varying inversely). If, on a given trial, Table 1 dictated cooperation, the computer's cooperation probability would move up by 25% (or, if at 100%, would stay there); if Table 1 dictated defection, the computer's cooperation probability would move down by 25% (or if at 0% would stay there).

One reason for applying the computer's strategy probabilistically instead of all-or-none was, similar to the reason for alternate trials, to better model everyday life social interactions—where strategies are seldom rigidly applied. Another reason had to do with the extreme sensitivity of the Pavlov strategy to initial conditions. For example, suppose a participant emitted a long string of cooperation choices. If the string of cooperations followed a cooperation by the computer, the computer would continue to cooperate and the participant would earn 5 U on each trial; however, if the string of cooperations followed a defection by the computer, the computer would continue to defect and the participant would earn only 1 U on each trial. Assuming that the participant's string of cooperation choices is as likely to follow a cooperation as a defection by the computer, the participant technically would earn an average of $(5+1)/2$ or 3 U per trial. Strings of defections against Pavlov, on the other hand, would cause the computer to alternate between cooperation and defection as first one then the other was punished—yielding 6 and 2 U on alternate trials, averaging 4 U per trial. Note that against Pavlov, as opposed to tit-for-tat, average theoretical earnings increase with defection. It is unlikely, however, that a participant would rigidly maintain a pattern regardless of the computer's responses or the participant's own earnings. That is, a string of cooperation choices leading to cooperation by the computer and 5 U per trial would be more likely maintained than one leading to defection by the computer and 1 U per trial. With rigid application of Pavlov, points of non-optimal equilibrium based on initial conditions could develop. The probabilistic procedure was instituted to break up such patterns.

2. Experiment 1

The purpose of Experiment 1 was to compare the effect on participants' choices of playing against tit-for-tat and Pavlov.

2.1. Method

2.1.1. Participants

Thirteen participants (six female, seven male) played against the tit-for-tat strategy and 13 participants (five female, eight male) played against the Pavlov strategy. Two participants (one male in the tit-for-tat group and one male in the Pavlov group) were excluded from the data analysis because they did not sample both choices at least five times each. Participants were given research credit that counted towards the fulfillment of a course requirement and $0.01 for each point they earned during the experiment ($6.00 maximum).

2.1.2. Procedure

Pentium computers with 17-inch monitors were used to run a Visual Basic program that displayed the game and recorded participants' responses. During the game, participants earned points by clicking the 'Left' command button (the equivalent of a cooperation choice) or the 'Right' command button (the equivalent of a defection choice) with the mouse (see Fig. 2). The game always began with the computer choosing 'White' (the equivalent of a cooperation choice) and on subsequent trials, the computer always chose before the participant. After the first trial, the computer chose 'White' (the equivalent of a cooperation choice) or 'Black' (the equivalent of a defection choice) according to a probability that could increase or decrease by 25% each trial (starting at 50%). The computer's probability of defection was the inverse of its probability of cooperation. For simplicity, the manner in which the probability changed will be discussed in terms of the computer choosing cooperation (White). The only difference between the tit-for-tat and Pavlov conditions was how the probability of cooperation changed from trial to trial.

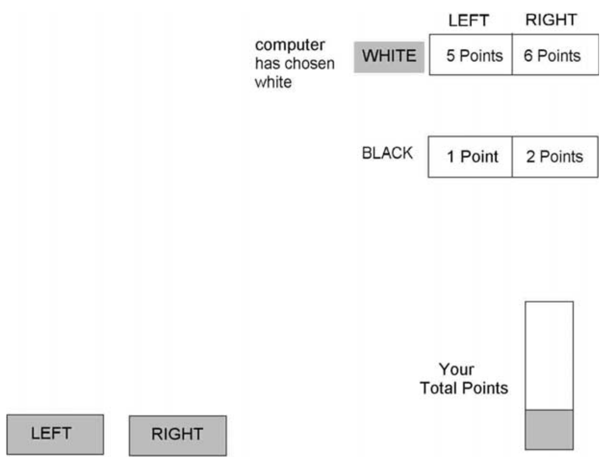

Fig. 2. Starting screen display in Experiment 1. The computer chose white (cooperation) or black (defection) probabilistically. On each trial, the probability changed according to the tit-for-tat or Pavlov contingencies. Participants earned points from the payoff matrix by clicking on the 'Left' command button (cooperation) or the 'Right' command button (defection).

During the tit-for-tat condition, the probability that the computer would cooperate changed each trial according to two rules (see Table 1): If the participant cooperated (Left), the probability that the computer would cooperate on the next trial increased by 25% (until it reached 100%); if the participant defected (Right), the probability that the computer would cooperate on the next trial decreased by 25% (until it reached 0%).

During the Pavlov condition, the probability that the computer would cooperate changed each trial according to four rules (see Table 1): If both the computer and the participant cooperated, then the probability increased by 25% (until it reached 100%); if the computer cooperated and the participant defected, then the probability decreased by 25% (until it reached 0%); if the computer defected and the participant cooperated, then the probability decreased by 25% (until it reached 0%); and if both the computer and the participant defected, then the probability increased by 25% (until it reached 100%).

One to six participants were run at a time and these groups were randomly assigned to either the tit-for-tat or Pavlov condition. Each participant was seated in front of a computer that displayed the screen shown in Fig. 2. The computer stations were separated by office partitions. The experimenter assured the participants that they would receive money at the end of the experiment (along with research credit) and then read instructions to the participants.

After reading the instructions, the experimenter walked behind the participants and checked their screens to make sure they started the game. The experimenter clarified the instructions to participants who were confused.

The game started after the 'Left' or 'Right' command button was clicked with the mouse and ended after 100 trials with a message that read: 'Thanks for participating—Trials complete Total Points = # #.' When the computer chose 'White,' a message that said: 'Computer has Chosen White' appeared next to the two upper boxes and the word 'White' was highlighted in blue. On these trials participants had a choice between earning five or six points. That is, a participant received five points for clicking 'Left' and six points for clicking 'Right.' When the computer chose 'Black,' a message that said: 'Computer has Chosen Black' appeared next to the two lower boxes and the word 'Black' was highlighted in blue. On these trials, participants had a choice between earning one or two points. That is, a participant received one point for clicking 'Left' and two points for clicking 'Right.'

After the computer picked a color, nothing happened until the participant clicked on the 'Left' or the 'Right' command button with the mouse. Immediately after a command button was clicked, both command buttons were disabled, a blue bar appeared under the command button that was clicked, the number of points earned on that trial was indicated by the appropriate number in the payoff matrix being highlighted in blue and the height of the bar in the total points box (colored green on the monitor) increased proportionally to the number of points earned on that trial. The screen remained frozen for 6 s; then the computer's choice on the next trial was indicated, as described above, and the command buttons were enabled. This completed the trial.

After all participants had finished playing the game, each participant was debriefed and paid $0.01 for each point earned during the game.

Participants who did not sample both choices at least five times each were not included in the data analysis.

2.2. Results

Fig. 3 shows the mean percentage of cooperation choices across subjects in the tit-for-tat and Pavlov groups for the first 20 and last 20 of 100 experimental trials. Both strategies increased the cooperation of participants playing against them. The mean percent cooperation during the last 20 trials was significantly higher than the mean percent cooperation during the first 20 trials across participants in both the tit-for-tat and Pavlov groups, $t(11) = 2.911$, $P = 0.014$ and $t(11) = 2.253$, $P = 0.046$, respectively.

Fig. 4 shows the distributions of cooperation over the last 20 trials for the two groups. Of the group playing against Pavlov, six players cooperated on all 20 trials and the rest were unsystemat-

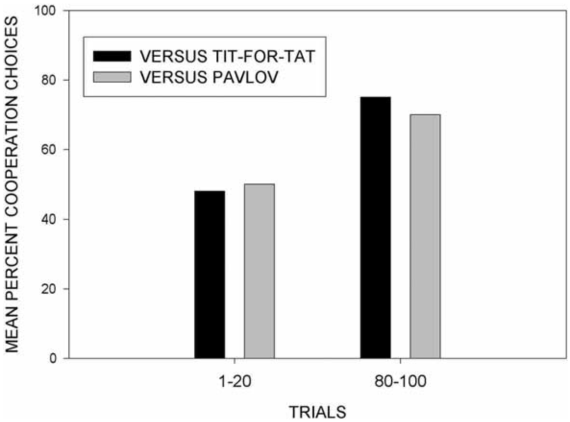

Fig. 3. Mean percentage of cooperation choices across subjects in the tit-for-tat and Pavlov groups for the first 20 and last 20 of the 100 experimental trials in Experiment 1.

ically distributed across the scale. Similarly, of the group playing against tit-for-tat, five players cooperated on all 20 trials and the rest were unsystematically distributed across the scale. The two groups did not significantly differ in the number of participants who cooperated on all of the last 20 trials, $\chi^2 = 0.17$, $P = 0.682$.

There was no significant difference in the average number of points earned during the last 20 trials for the tit-for-tat ($M = 86.9$) and Pavlov ($M = 94.9$) groups, $t = 1.042$, $P = 0.309$.

2.3. Discussion

The fact that cooperation by participants in both the Pavlov and tit-for-tat groups increased over the course of the experiment indicates that participants in both groups must have been sensitive to feedback, learning to cooperate when initial conditions were such as to reinforce cooperation (when both the computer and the subject had cooperated on the previous trial) and not under other conditions. All six of the tit-for-tat participants and all five of the Pavlov participants who cooperated on all of the last 20 trials earned five points on each of those trials.

The distributions of Fig. 4 are far from normal; they indicate that some participants in each group cooperated over the final 20 (of 100) trials, while the remaining participants showed no consistent pattern. Those participants, in either group, who learned anything in this experiment, learned to cooperate. This was somewhat surprising for the Pavlov group, since defection was not resolutely punished against Pavlov as it was against tit-for-tat. Consistent defection against tit-for-tat would earn two points per trial while, against Pavlov, consistent defection would earn four points per trial (six and two points alternately). The steady five points per trial earned by consistent cooperation might not have been expected to dominate the average four points earned for consistent defection in the Pavlov group, especially since the participant would always earn one more point by defecting than by cooperating on the current trial.

The next experiment was designed to further compare these probabilistic versions of tit-for-tat and Pavlov in two contexts different from that of Experiment 1. One context provided a signal (a spinner) specifically indicating the computer's probability of cooperating on each trial. The other context implied to the participant that the computer's choices were being made by another player. These different contexts were expected to interact with the probabilistic tit-for-tat and Pavlov contingencies to produce different patterns of cooperation.

3. Experiment 2

The purpose of Experiment 2 was to compare the effects of playing against tit-for-tat and Pavlov in different contexts—one that emphasized the automatic nature of the computer's play and one that implied that the participant was playing against a human being rather than a computer.

3.1. Method

3.1.1. Participants

Forty-four undergraduate students (24 female, 20 male) attending the State University of New York at Stony Brook participated in the experiment. Twenty-four participants (14 versus tit-for-tat, ten versus Pavlov) played in the spinner context—a visible spinner indicated the comput-

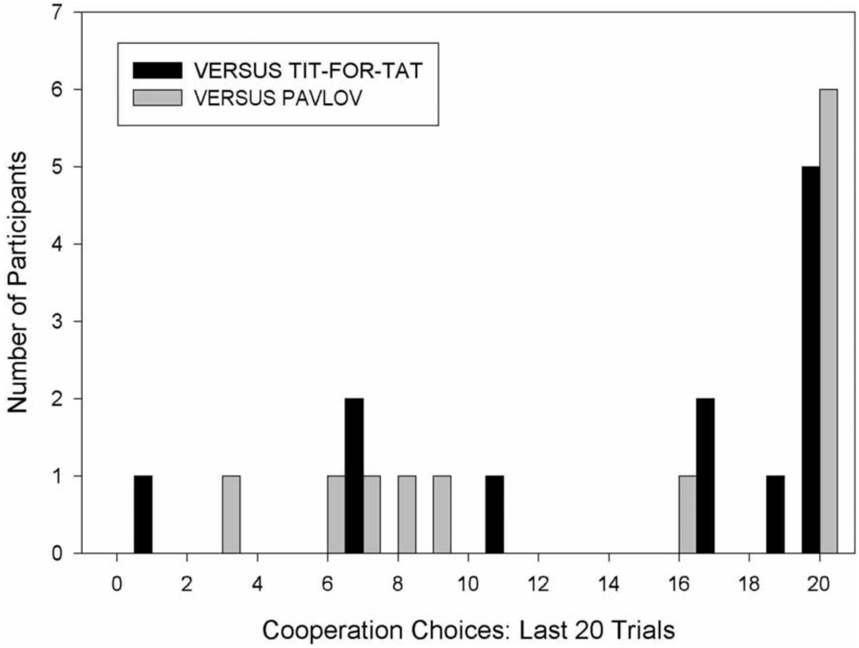

Fig. 4. Number of cooperation choices by participants over the last 20 trials for the tit-for-tat and Pavlov groups in Experiment 1.

er's probability of cooperation. Two participants (both male) both playing versus Pavlov were excluded from the data analysis because they did not sample both choices at least five times each. Twenty participants (ten versus tit-for-tat, ten versus Pavlov) played in the 'other player' context—without the spinner but with instructions implying that they were playing, not against a computer, but against another of the participants playing at the same time. Two participants in the 'other player' group (one male, one female), both playing against tit-for-tat, were excluded from the data analysis because they did not sample both choices at least five times each.

3.1.2. Procedure

The tit-for-tat and Pavlov games of Experiment 2 were the same as those of Experiment 1 except that their contexts differed. In the spinner context, a spinner with four 90° sectors was added to the upper left corner of the display shown in Fig. 2. The proportion of white indicated the computer's probability of cooperation. As in Experiment 1, at the end of each trial the screen showing points earned and the computer's and participant's choices on that trial, remained as it was for 2 s and then, according the game's rules, one of the four pie pieces in the wheel could be changed from white to black or from black to white. After updating the wheel, depending on the previous trial's choices, the screen remained frozen for another 4 s and then the wheel spun for ≈ 1.5 s. The percentage of white on a wheel reflected the current probability that the wheel would stop on white; vice-versa for black. For example, there was a 100% probability that the computer would choose 'White' when the wheel was totally white and there was a 0% probability that the computer would choose 'White' when the wheel was totally black. After the wheel stopped spinning, the word 'White' or 'Black' was highlighted in blue and the appropriate message appeared next to it. For example, if the wheel landed on black, then the computer chose 'Black' and a participant had a choice between earning one or two points on that trial. After the computer chose a color, the blue bar under the command button that was chosen disappeared, the blue highlight on the number in the payoff matrix disappeared and the command buttons were enabled again. This completed the trial.

In the ‘other player’ context, the tit-for-tat and Pavlov games again operated the same way as those of Experiment 1 except that the screen was changed in order to convince participants that they were actually playing with another participant in the room. No spinner was displayed. Instead, a second payoff matrix, that of the other player, was displayed to the left of the participant's own payoff matrix. During the game, when it was the participant's turn, only the ‘Left’ and ‘Right’ command buttons were enabled; the right payoff matrix was framed by a box that contained the caption ‘Your Turn.’ The delay that occurred after a participant chose ‘Left’ or ‘Right’ lasted for only 3 s (instead of 6 s in Experiment 1 and for the spinners group of Experiment 2) and after this delay the frame around the right payoff matrix disappeared and the ‘other player's’ turn began. There was actually no ‘other player’; instead the computer determined the ‘other player's’ choices.

The ‘other player's’ turn was exactly like the participant's turn except for the following. The ‘other player's’ turn began with the left payoff matrix framed by a box that contained the caption ‘Other Player's Turn.’ Only the ‘Top’ and ‘Bottom’ command buttons were enabled, but nothing happened if the participant clicked on them. The computer, according to the tit-for-tat or Pavlov contingencies, chose ‘Top’ (the equivalent of a cooperation choice) or ‘Bottom’ (the equivalent of a defection choice) 1.5 s, on average, after the start of its turn. A blue bar appeared under the ‘other player's’ choice (‘Top’ or ‘Bottom’ command button) and the number of points that the ‘other player’ earned was highlighted in the left payoff matrix. The average amount of time between a choice made by a participant and the start of the next trial was programmed to be roughly the same as it was for other groups (6 s). The frame around the left payoff matrix disappeared on completion of the ‘other player's’ turn.

Groups of four or six participants were run at a time and these groups were quasi-randomly assigned to either the tit-for-tat or Pavlov group. The instructions read to participants were similar to those of Experiment 1; however, participants in the ‘other player’ group were told that they were going to play the game with another participant in the room and they would never find out with whom they played. After the game, each participant in the ‘other player’ group was asked to write at least two sentences for each of the following questions: (1) ‘What did you think of the game?’ (2) ‘What did you think of the other player?’ These questions were used only to ascertain whether the participants actually believed they were playing another person. None of the participants expressed any doubt as to whether they were actually playing with another person.

3.2. Results

Fig. 5(a) shows the mean percentage of cooperation choices across participants playing in the spinners context versus tit-for-tat and Pavlov for the first five and last five of 100 experimental trials. Tit-for-tat succeeded in increasing the cooperation of participants playing against it while Pavlov did not. This is evidenced by the fact that the mean percent cooperation during the last 20 trials was significantly higher than the mean percent cooperation during the first 20 trials across participants in the tit-for-tat group, but not in the Pavlov group; $t(9) = 3.537$, $P = 0.006$ and $t(11) = 0.386$, $P = 0.707$, respectively.

Fig. 5(b) shows the mean percentage of cooperation choices across participants playing in the ‘other player’ context versus tit-for-tat and Pavlov for the first 20 and last 20 of the 100 experimental trials. The effects of both strategies on cooperation were diminished in the social context. The mean percent cooperation during the last 20 trials was not significantly higher than the mean percent cooperation during the first 20 trials across participants versus tit-for-tat, $t(8) = 0.199$, $P = 0.848$; but, although not statistically significant, this trend was evident in the Pavlov group, $t(9) = 1.739$, $P = 0.116$.

Fig. 6(a) shows the distributions of cooperation over the last 20 trials for participants playing in the spinner context versus tit-for-tat and Pavlov. Versus Pavlov, one player cooperated on all 20

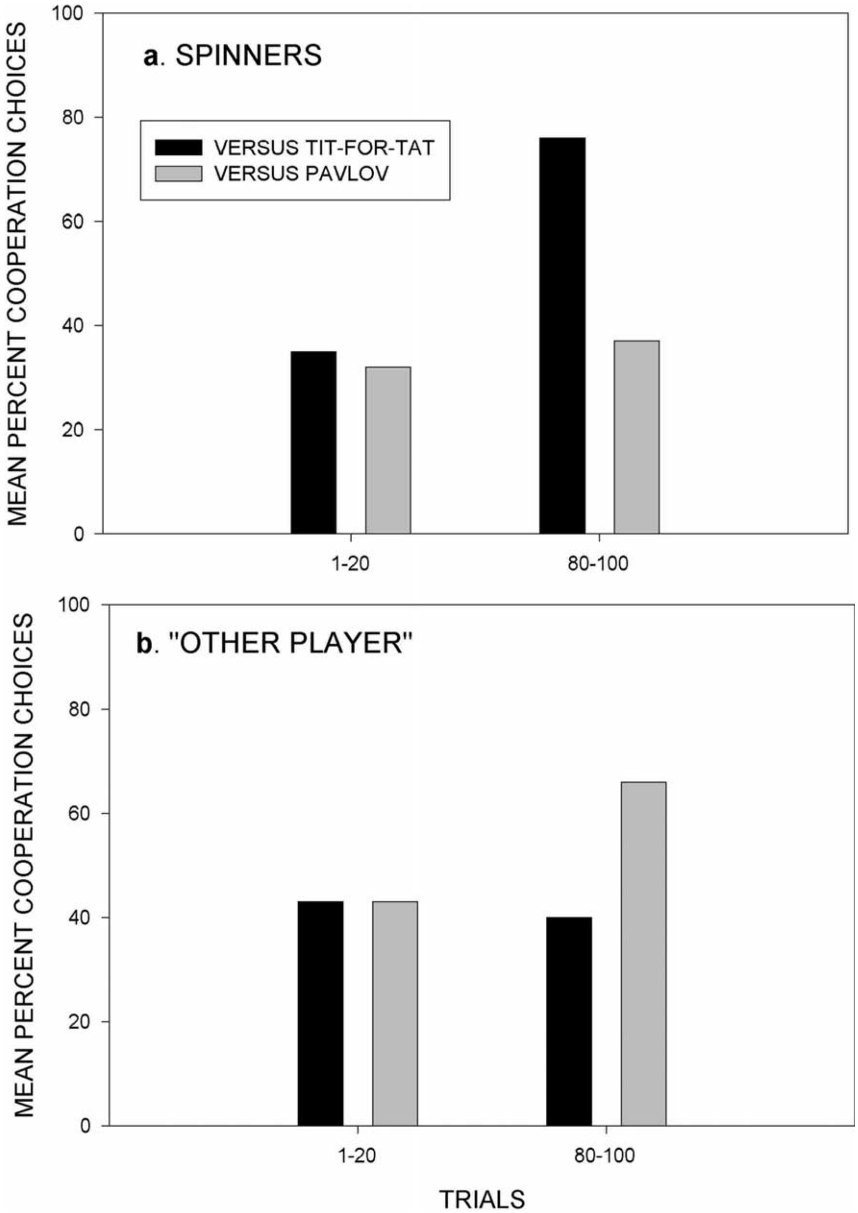

Fig. 5. Mean percentage of cooperation choices across subjects in the tit-for-tat and Pavlov groups for the first 20 and last 20 of the 100 experimental trials in the spinner context (upper graph) and the ‘other player’ context (lower graph).

trials, one defected on all 20 trials and the rest were unsystematically distributed across the scale. Versus tit-for-tat, five players cooperated on all 20 trials while five were closely grouped around co-operation on half of the trials.

Fig. 6(b) shows the distributions of cooperation over the last 20 trials for participants playing in the ‘other player’ context against tit-for-tat and Pavlov. Versus Pavlov, five participants cooperated on all 20 trials, two defected on all 20 trials and the rest were unsystematically distributed across the scale. Versus tit-for-tat, only one par-

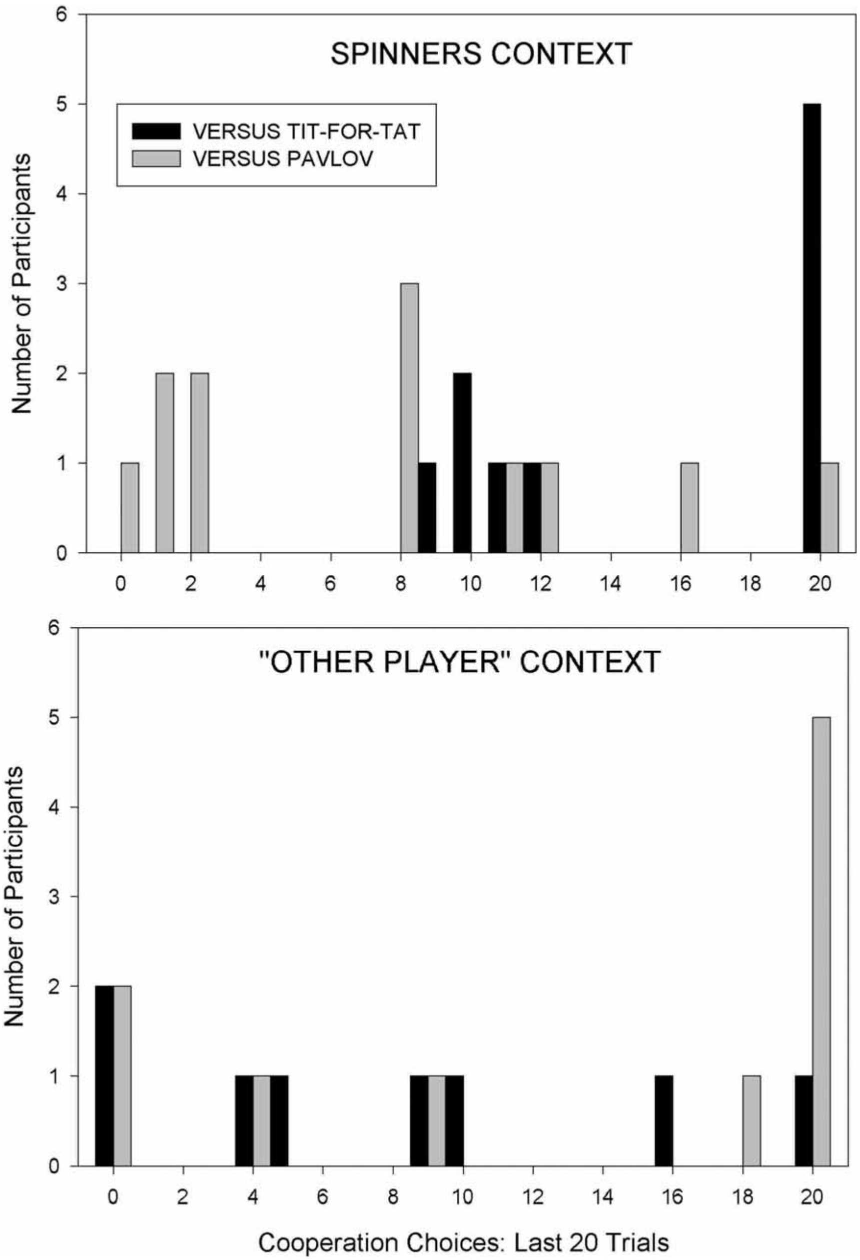

Fig. 6. Number of cooperation choices by participants over the last 20 trials for the tit-for-tat and Pavlov groups in Experiment 2.

ticipant cooperated on all 20 trials, two defected on all 20 trials and the rest were unsystematically distributed across the scale. As in Experiment 1, the distributions are highly skewed; again, some participants consistently cooperated over the last 20 trials, while the remainder were scattered unsystematically across the distribution. Taking the number of participants cooperating on all of the last 20 trials as opposed to those not cooperating on all of the last 20 trials, there was a significant interaction on this measure between context (spinner versus 'other player') and computer strategy (tit-for-tat versus Pavlov; log linear analysis, $\chi^2 = 6.8$, $P = 0.009$). That is, whether or not a participant cooperated on all of the last 20 trials depended on both computer strategy and context and did not depend on either computer strategy or context alone. Participants were more likely to cooperate on the last 20 trials in the spinners context against tit-for-tat and in the 'other player' context against Pavlov than in the spinners context against Pavlov and in the 'other player' context against tit-for-tat.

Earnings over the last 20 trials showed the same pattern as cooperations. In the spinners context, participants playing against tit-for-tat earned more points on average over the last 20 trials ($M = 96.4$) than those playing against Pavlov ($M = 85.9$), $t = 2.058$, $P = 0.0529$. In the 'other player' context, on the other hand, participants playing against Pavlov earned more points on average over the last 20 trials ($M = 94.1$) than those playing against tit-for-tat ($M = 61.5$), $t = 3.603$, $P = 0.002$; further evidence that performance in the two contexts was differentially affected by the two strategies.

Fig. 7 shows, for each participant in both experiments, average earnings per trial as a function of number of cooperations over the last 20 trials. The upper graph shows earnings as a function of cooperations against tit-for-tat; the lower graph, against Pavlov. The numbers next to some points indicate number of overlaps. These functions are obtained 'feedback functions;' they are empirical versions of the theoretical feedback function in Fig. 1(b). The variability is due to both the probabilistic nature of the computer's strategy and, especially versus Pavlov, the strong dependence of the computer's choices on initial conditions.

Note the steepness of the tit-for-tat feedback function relative to the Pavlov feedback function. The difference in steepness lies almost wholly in the left half of the graphs, reflecting the fact that repeated defection against tit-for-tat results in repeated defection by the computer, earning the participant only two points per trial, whereas repeated defection against Pavlov results in alternation between cooperation and defection by the computer, earning the participant six and two points on alternate trials (see Table 1 and Fig. 1a).

Fig. 7 shows that the theoretical maximum reinforcement rate (five points per trial) was approached or exceeded, with both Pavlov and tit-for-tat contingencies, over a range between ten and 20 cooperation choices over the last 20 trials. This occurred because the participant's cooperation on $< 100\%$ of the trials could earn an average of five points per trial against either computer strategy—a player could cooperate enough to drive the computer's probability of cooperation to 100% and then occasionally defect, reducing the computer's next-trial probability of cooperation to 75% rather than 0% as in non-probabilistic prisoner's dilemma games. The difference between this tactic and repeated cooperation is that number of points earned by players who cooperated on between 50 and 100% of the trials varied and could be more or less than five per trial while, with continuous cooperation, number of points earned was steady at five per trial.

Cooperation at a rate between 50 and 100% in the present experiment was therefore tantamount to gambling. A participant's tendency to avoid risk would have acted to produce repeated cooperation. (Countering this tendency, however, was the immediate reward increase always available for defecting.)

3.3. Discussion

About half of the participants in Experiment 1, with no specific signal indicating the computer's probability of cooperation and with no belief that they were playing against another player, learned to cooperate against both tit-for-tat and Pavlov strategies. On the other hand, in Experiment 2, a specific signal indicating the computer's probability of cooperation and its dependence on the participant's choices hindered learning of cooperation against Pavlov, whereas the participants' belief that they were playing against another player rather than a computer hindered learning of cooperation against tit-for-tat.

One explanation for this pattern of results is that (along with immediate reinforcement for de-

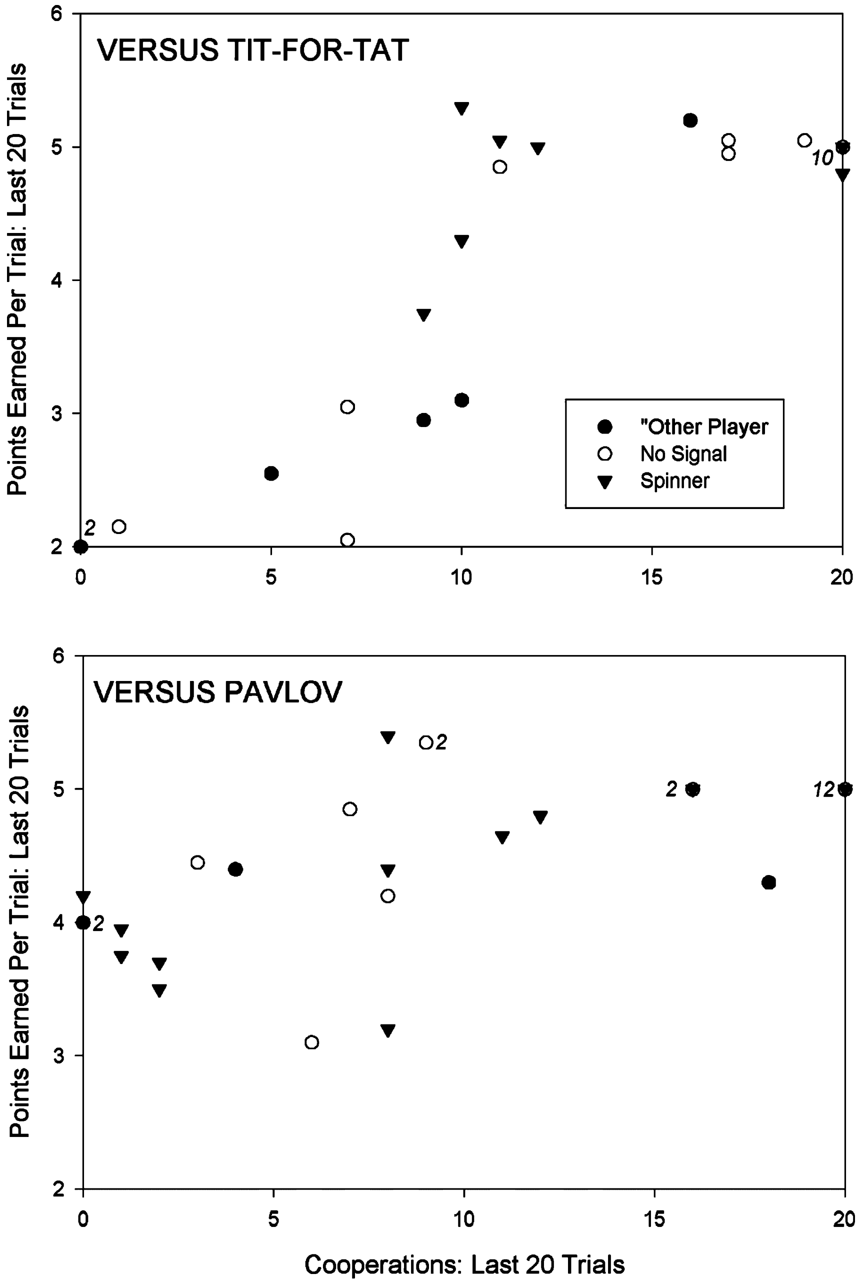

Fig. 7. Average points earned per trial over the last 20 trials for each participant in Experiments 1 and 2.

fection present in all prisoner's dilemma games) a pattern of interaction with the feedback contingencies may form a unit of behavior subject to reinforcement and punishment. Two such patterns, corresponding to the two computer strategies, are (a) a tendency to adapt to the computer's play—to learn from the computer; and (b) a tendency to modify the computer's play—to teach the computer. With the unsignaled conditions of Experiment 1, participants may have begun the experiment without bias towards one or the other of these strategies. Although participants knew that they were playing against a machine in Experiment 1, they did not know how the machine made its decisions. Against tit-for-tat, a tendency to learn from the computer (to play Pavlov against tit-for-tat) would have led to mutual cooperation and five points per trial while a tendency to teach the computer (to play tit-for-tat against tit-for-tat) would, if participants were more inclined to defect than to cooperate (defection being immediately reinforced), have led to mutual defection and two points per trial.

On the other hand, against Pavlov, a tendency to learn from the computer (to play Pavlov against Pavlov) would have led to a complex, unstable pattern strongly dependent on chance and, in any case, less efficient than tit-for-tat.[1] On the other hand, a tendency to teach the computer (to play tit-for-tat against Pavlov) would have led to mutual cooperation and five points per trial. Those participants in Experiment 1 who learned to cooperate could have tried out strategies in turn, settling, in general, on the one that seemed to work — teaching (playing tit-for-tat) against Pavlov, learning (playing Pavlov) against tit-for-tat.

The two contexts of Experiment 2, however, may have created biases which persisted despite less than optimal results. The 'other player' context may have biased participants to adopt a teaching strategy. In the 'other player' context, the points supposedly earned by the 'other player' were clearly displayed while the computer's automatic response was obscured by the absence of the spinner. The participants believed that their opponents were human beings and therefore probably sensitive to reinforcement and punishment. Against Pavlov, persistence by the participant of this teaching strategy would have soon led to mutual cooperation; against tit-for-tat, to mutual defection.

[1] Playing Pavlov against the normal, non-probabilistic Pavlov strategy would lead to a repeated three-trial pattern, two defects and one cooperate (earning six, one and two points), averaging three points per trial.

The spinners context provided a veridical indication of the computer's response to the participants' choices more reliable than points earned. For example, a cooperation choice against tit-for tat would have almost immediately changed a 50% white spinner to a 75% white spinner. If, on the next trial, the spinner then landed on black anyway, a participant playing against tit-for-tat with spinners visible would have seen the immediate increase in the computer's cooperation probability whereas, without spinners visible, a participant would have seen only the computer's eventual defection. This immediate feedback could have brought choice under the control of the non-probabilistic feedback functions of Fig. 1(b) rather than the probabilistic ones of Fig. 7. Against tit-for-tat such sharpening of control might have made little difference, since in both cases strings of defections were severely punished. However, against Pavlov, repeated defection was reinforced on alternate trials by the computer's increased probability of cooperation (visible on the spinner). This conditioned reward together with the immediate reward for defection always present with prisoner's dilemma contingencies could explain why participants playing against Pavlov in the spinners context did not learn to cooperate.

In the spinners context, moreover, the participants knew that they would have to adapt to the machine's behavior as represented by the displayed spinner. Against tit-for-tat, persistence by the participant of this learning strategy would have led to mutual cooperation; against Pavlov, to a result (an average of three points per trial) worse than that of repeated defection (an average of four points per trial). Participants in this group thus may have given up on any strategy and simply defected on most trials (as indicated by the cluster of filled-in triangles on the lower graph of Fig. 7). The veridical information provided by the spinner actually hindered the participants playing against Pavlov — they did not expect that a spinner's behavior could be reinforced or punished and did not try to do so.

4. Conclusions

Playing a prisoner's dilemma game against either tit-for-tat or Pavlov strategies is a self-control task in that the alternatives are an immediate smaller reward versus a delayed larger reward. But the prisoner's dilemma does not present these alternatives as a simple choice on any given trial, as many laboratory studies of self-control do (Logue, 1988). Rather, in the prisoner's dilemma, the consequences of current-trial choice cannot be realized, even with a delay, on the current trial. To obtain the larger reward, the prisoner's dilemma player must actually choose the smaller reward on the present trial (cooperate on the present trial), obtain that reward and wait for any intertrial interval. Only then is the larger reward obtained. Each choice thus serves two purposes—it obtains the reward earned by the previous trial's choice and it affects the alternatives to be presented on the next trial.

In complex self-control tasks, where choices and consequences overlap, reinforcement may be maximized by the establishment of behavioral patterns extending across multiple choices (Kudadjie-Gyamfi and Rachlin, 1996). These self-generated patterns are reinforced only after they are formed (Rachlin, 1995, 2000). In some cases, as in the present experiment, the reinforcement (points exchangeable for money) is extrinsic to the patterns but, in many cases in everyday life, reinforcement is intrinsic in the patterns themselves (as in learning to play a musical instrument or to dance or to play a complex sport). Once established, however, such patterns may be maintained by the higher reinforcement rate they themselves generate. Outside of the laboratory, various patterns would be appropriate in various situations and thus come under stimulus control. The two contexts of Experiment 2—playing against a machine and playing against another person—may have each controlled a different pattern. Participants playing against a machine that made its decisions automatically, by spinning a wheel, may have been strongly biased to adapt to the machine's programmed contingencies while participants who believed that they were playing against another human being may have been strongly biased to alter the 'other player's' behavior by reinforcing or punishing his or her choices. These two biases could have interacted with the machine's own strategies to produce the differing performances found in Experiment 2.

Biological theories of altruistic behavior based on an analogy between prisoner's dilemma and everyday social interaction frequently characterize individuals as biased to cooperate (altruistic) or biased to defect (selfish). But, as the present experiments indicate, this classification may be too simple. In everyday life, as in these experiments, people approach machines and other people, not with simple tendencies to be altruistic or selfish, but with different behavioral patterns, different strategies of interaction. Perhaps this is why we are offended when we expect to be dealing with another person (as when receiving a telephone call) and discover that we are dealing with a machine or, more vividly, when the people we deal with act like machines ('This is our policy sir.')—insensitive to our practiced strategies of behavioral control.

Acknowledgements

The research was supported by grants to the second author from the National Institute of Mental Health and the National Institute for Drug Abuse. We thank James C. Baker for programming the experiments.

References

Ainslie, G., 1992. Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person. Cambridge University Press, New York.

Axelrod, R., 1984. The Evolution of Cooperation. Basic Books, New York.

Caporael, L.R., Dawes, R.M., Orbell, J.M., van de Kragt, A.J.C., 1989. Selfishness examined: cooperation in the absence of egoistic incentives. Behavioral and Brain Sciences 12, 683–739.

Dawes, R., 1980. Social dilemmas. Annual Review of Psychology 31, 169–193.

Fudenberg, D., Maskin, E., 1990. Evolution and cooperation in noisy repeated games. New Developments in Economic Theory 80, 274–279.

Green, L., Price, P.C., Hamburger, M.E., 1995. Prisoner's dilemma and the pigeon: control by immediate consequences. Journal of the Experimental Analysis of Behavior 64, 1–17.

Kollock, P., 1998. Social dilemmas: the anatomy of cooperation. Annual Review of Sociology 24, 183–214.

Komorita, S.S., 1994. Social Dilemmas. Brown Communications, Inc, Dubuque, IA.

Komorita, S.S., Parks, C.D., 1995. Interpersonal relations: mixed-motive interaction. Annual Review of Psychology 46, 183–207.

Komorita, S.S., Hilty, J.A., Parks, C.D., 1991. Reciprocity and cooperation in social dilemmas. Journal of Conflict Resolution 35, 494–518.

Kraines, D., Kraines, V., 1989. Pavlov and the prisoner's dilemma. Theory and Decision 26, 47–79.

Kraines, D., Kraines, V., 1993. Learning to cooperate with Pavlov: an adaptive strategy for the prisoner's dilemma with noise. Theory and Decision 35, 107–150.

Kraines, D., Kraines, V., 1995. Evolution of learning among Pavlov strategies in a competitive environment with noise. Journal of Conflict Resolution 39, 439–466.

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

Logue, A.W., 1988. Research on self-control: an integrating framework. Behavioral and Brain Sciences 11, 665–679.

Macy, M.W., 1995. Pavlov and the evolution of cooperation: an experimental test. Social Psychology Quarterly 58, 74–87.

Messick, D.M., Brewer, M.B., 1983. Solving social dilemmas. In: Wheeler, L., Shaver, P. (Eds.), Review of Personality and Social Psychology. Sage, Beverly Hills, CA, pp. 11–44.

Messick, D.M., Liebrand, W.B.G., 1995. Individual heuristics and the dynamics of cooperation in large groups. Psychological Review 102, 131–145.

Nowak, M., Sigmund, K., 1993. A strategy of win-stay-lose shift that outperforms tit-for-tat in the prisoner's dilemma game. Nature 364, 56–58.

Rachlin, H., 1995. Self-control: beyond commitment. Behavioral and Brain Sciences 18, 109–159.

Rachlin, H., 2000. The Science of Self-Control. Harvard University Press, Cambridge, MA.

Rachlin, H., Brown, J., Baker, F., 2000. Reinforcement and punishment in the prisoner's dilemma game. In: Medin, D. (Ed.), The Psychology of Learning and Motivation, vol. 40. Academic Press, New York.

Rapoport, A., 1967. Optimal policies for the prisoner's dilemma. Psychological Review 74, 136–148.

Rapoport, A., Chammah, A., 1965. Prisoner's Dilemma. University of Michigan Press, Ann Arbor.

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

Stephens, D.W., 2000. Cumulative benefit games: achieving cooperation when players discount the future. Journal of Theoretical Biology 205, 1–16.

Van Lange, P.A.M., Liebrand, B.G., Messick, D.M., Wilke, H.A.M., 1992. Social Dilemmas: Theoretical Issues and Research Findings. Pergamon, Oxford.