Learning & Behavior
2003, 31 (4), 318-331

Learning by pigeons playing against tit-for-tat in an operant prisoner's dilemma

FEDERICO SANABRIA
State University of New York, Stony Brook, New York

FOREST BAKER
Illinois Wesleyan University, Bloomington, Illinois

and

HOWARD RACHLIN
State University of New York, Stony Brook, New York

Each of four pigeons was exposed to a single random-ratio schedule of reinforcement in which the probability of reinforcement for a peck on either of two keys was 1/25. Reinforcer amounts were determined by an iterated prisoner's dilemma (IPD) matrix in which the "other player" (a computer) played *tit-for-tat*. One key served as the *cooperation* (C) key; the other served as the *defection* (D) key. If a peck was scheduled to be reinforced and the D-key was pecked, the immediate reinforcer of that peck was always higher than it would have been had the C-key been pecked. However, if the C-key was pecked and the *following* peck was scheduled to be reinforced, reinforcement amount for pecks on either key were higher than they would have been if the previous peck had been on the D-key. Although immediate reinforcement was always higher for D-pecks, the overall reinforcement rate increased linearly with the proportion of C-pecks. C-pecks thus constituted a form of self-control. All the pigeons initially defected with this procedure. However, when feedback signals were introduced that indicated which key had last been pecked, *cooperation* (relative rate of C-pecks)—hence, self-control—increased for all the pigeons.

In traditional self-control studies, subjects choose between a smaller reward that is delivered sooner and a larger reward that is delivered later (e.g., Green, Fristoe, & Myerson, 1994; Mazur, 1987; Rachlin & Green, 1972). Such a paradigm does not directly model many natural situations of self-control. In natural situations, larger and smaller rewards are often simultaneously available; choosing the larger alternative maximizes the local reinforcement rate but reduces overall, or global, value. (For example, having a second dessert may be immediately worth more than not having it but may contribute to bad health in the long run.) Choosing the smaller alternative (refusing the dessert) may have the reverse effect. Research on self-control has increasingly adopted this local versus global paradigm (e.g., Green, Price, & Hamburger, 1995; Herrnstein, Loewenstein, Prelec, & Vaughan, 1993; Heyman & Tanz, 1995; Rachlin, Brown, & Baker, 2001) and has used it as a model of addictive behavior (Herrnstein & Prelec, 1992; Rachlin, 2000). A relatively simple version of this paradigm is an iterated prisoner's dilemma (IPD), in which subjects play against a reciprocal strategy called *tit-for-tat*. Before discussing the tit-for-tat strategy, let us consider the contingencies of a two-player PD in which neither player plays with a fixed strategy.

In this game, two players each have two choice alternatives (*cooperating* and *defecting*). Each of the four choice combinations is associated with a different outcome (reward or punishment magnitude). The PD payoff matrix (Figure 1) shows how the outcome for each player is based on the choices of both players (examples of reward sizes for each player are shown in parentheses).

Note that defection pays more than cooperation on any given trial for either player, regardless of what the other player chooses. If Player A cooperates, Player B should defect and obtain the "best" outcome instead of settling for the "good" outcome; if Player A defects, Player B should also defect to obtain the "bad" outcome and avoid getting the "worst" outcome. The same contingencies apply for Player A. Thus, in a one-shot PD (where the players choose an action only once—i.e., in one trial), each player would maximize reinforcement by defecting. If both players maximized reinforcement, both would defect; each would earn the "bad" reward. However, if both players chose the lesser reward, both would cooperate; each would earn the "good" reward. This conflict between reward to the individual (obtained by defection) and reward to the group (obtained by cooperation) creates the dilemma.

The most important difference between an IPD (in which the players choose repeatedly) and a one-shot PD

This research was supported by a grant from the National Institute for Mental Health. Correspondence concerning this article should be addressed to F. Sanabria, Department of Psychology, State University of New York, Stony Brook, NY 11794-2500 (e-mail: federico.sanabria@sunysb.edu).

		Player A: Cooperates	Player A: Defects
Player B: Cooperates		Good (5), Good (5)	Worst (2), Best (6)
Player B: Defects		Best (6), Worst (2)	Bad (3), Bad (3)

Figure 1. Prisoner's dilemma payoff matrix.

is that, in an IPD, it is possible to distinguish the immediate short-term preference for an action from the long-term preference for a sequence of actions (a strategy). As in a one-shot PD, defection is preferred in the short term because it always pays to defect on the current trial. The long-term payoff, on the other hand, depends on how the action of one player on one trial influences the actions of the other player on future trials. For instance, it is in the best (long-term) interest of a player to cooperate if that cooperation increases the probability that the other player will cooperate too.

In an IPD, each player's outcome is determined by that player's current choice and the other player's most recent choice. Each player obtains a high payoff if the other player cooperates (if B cooperates, A obtains either the "best" or the "good" outcome) and a low payoff if the other player defects (if B defects, A obtains either the "bad" or the "worst" outcome). Each player's cooperation rewards the other player and, in a repeated game, reinforces the other player's most recent choice. Correspondingly, each player's defection punishes the other player and, in a repeated game, punishes the other player's most recent choice (Rachlin et al., 2001). A strategy that makes use of these reward and punishment contingencies is called *tit-for-tat* (Axelrod, 1984). The tit-for-tat player cooperates on the first trial. After that, if the other player cooperates on a given trial, the tit-for-tat player will cooperate on the next trial; if the other player defects on a given trial, the tit-for-tat player will defect on the next trial. Thus, tit-for-tat reinforces cooperation and punishes defection.

With the numbers in Figure 1, assume that B plays tit-for-tat. In that case, if A always chose the higher (immediately preferred) reward, B would reciprocate, and A would obtain 3 reward units per trial in the future; however, if A always chose the lower (immediately dispreferred) reward, B would reciprocate, and A would obtain 5 reward units per trial in the future. Thus, in an IPD, against a tit-for-tat strategy, subjects must trade immediate (or local) reward for long-term (or global) reward. Rachlin (2000) calls this kind of choice, "complex ambivalence."[1]

In the PD literature, the labels *cooperation* and *defection* are used to describe the alternatives even when a single participant plays the game against a fixed strategy (Komorita, 1994; Rapoport & Chammah, 1965) and even when two computers play the game against each other (Axelrod, 1984). We retain this convention here, with the understanding that the terms have no larger meaning than that given by the matrix of Figure 1.

Previous studies of the IPD have shown that most human subjects eventually learn to cooperate against a tit-for-tat strategy (Baker & Rachlin, 2001; Komorita, 1994; Rachlin et al., 2001; Rapoport & Chammah, 1965; Silverstein, Cross, Brown, & Rachlin, 1998). Studies in which nonhuman subjects have been used, on the other hand, usually have not shown such learning (Green et al., 1995; Reboreda & Kacelnik, 1993). In Green et al.'s (1995) experiment, individual pigeons chose between pecking one key or another, one corresponding to cooperation and the other to defection. These responses were reinforced according to an IPD matrix in which a computer played tit-for-tat. Under these contingencies, pigeons generally defected.

Stephens, McLinn, and Stevens (2002) studied blue jays playing an IPD against a tit-for-tat strategy. The blue jays cooperated or defected by flying to a perch next to or away from a "stooge" blue jay that played tit-for-tat (by following a lit key). Cooperation was maintained (not acquired) only when reinforcers were accumulated over four-trial sequences, thereby eliminating the immediate advantage of defection. The study showed that blue jays are sensitive to delayed reinforcer magnitude in a highly interesting seminatural situation (they prefer a higher to a lower, equally delayed, reinforcer magnitude). However, without immediate reinforcement of defection, the blue jays faced no choice dilemma, either of social cooperation or of self-control.

Baker and Rachlin (2002) argued that the pigeons in Green et al.'s (1995) experiment might have defected because of the relatively long duration of each trial (25 sec). If a pigeon's choices were to have been affected by Green et al.'s (1995) experimental contingencies, its responses would have had to be sensitive to reinforcement delivered 32 sec later.[2] Accordingly, Baker and Rachlin (2002) attempted to increase cooperation by reducing the duration of each trial and providing feedback that signaled the previous trial's choice.

In Baker and Rachlin's (2002) trial-by-trial paradigm, pigeons chose between cooperation and defection by pecking 10 times on one of two keys. The signal on each trial depended on the pigeon's choice in the preceding trial. Each trial ended in reinforcement. Thus, three events intervened between a choice and its corresponding feedback signal: the feedback signal for the previous choice,

consumption of the reinforcer, and the next choice. Both the trial duration and the complex relationship between signal and response could have been responsible for the "room for improvement" left by this experiment (only 1 of 6 pigeons consistently cooperated on more than 80% of the trials with the shortest intertrial interval).

The main differences between the present free-operant procedure and Baker and Rachlin's (2002) trial-by-trial procedure were that, in the present experiment, each choice was a single peck and only 1/25th of the pecks were reinforced. With the free-operant procedure, feedback signals may follow (unreinforced) choice responses without an intervening reinforcement. This makes the just-previous choice more salient at the current moment than it would be in a trial-by-trial procedure. Against tit-for-tat, cooperation on the just-previous choice sharply elevates reward magnitude on the current choice. Thus, feedback signals indicating the just-previous choice may be expected to increase cooperation more in a free-operant procedure than they do in a trial-by-trial procedure. The purpose of the present experiment was to investigate this possibility.

The experiment was conducted in three phases. The aim of the first phase was to establish a baseline for cooperation (C-key pecks) in a free-operant procedure without signals and to compare it with cooperation after signals had been introduced. Once the level of cooperation with signals was determined, the aim of the second phase was to evaluate the degree of control of cooperation by the signals and the key locations. In separate conditions, the locations of C-pecks and D-pecks were reversed, and signals were eliminated. Finally, a third phase was designed to determine the effect of extended and repeated exposure to the signaled contingencies on the proportion of C-pecks.

METHOD

Subjects

Four experimentally naive male White Carneaux pigeons (Palmetto Pigeon Plant), maintained at approximately 80% of their free-feeding weights, were housed individually in a colony room, where they had free access to grit and water.

Apparatus

Sessions were conducted in four standard MED Associates modular test chambers (30.5 cm long, 24.1 cm wide, and 29.2 cm high), each enclosed in a sound- and light-attenuating box equipped with a ventilating fan. The front panel of each chamber contained three response keys arranged horizontally, each 2.54 cm in diameter, with their centers located 7 cm from the ceiling and separated from each other by 16.5 cm. All three keys could be illuminated from behind by colored (white, green, or red) light. The central key was used only during the training condition and remained dark and not functional for the rest of the experiment.

Food reinforcement was access to mixed grain delivered from a hopper located in the center of the front panel, 2 cm from the grid floor. During reinforcement, the hopper was illuminated with white light, and the keylights were extinguished. The chamber's houselight was never illuminated. A computer arranged experimental events and recorded data, using Med-PC for Windows.

Procedure

Training

Each subject was trained to peck on the central response key to obtain food from the hopper by reinforcing progressive approximations to keypecks (manual shaping). During training sessions, only the white key color was used. Training sessions ended after the subjects consistently obtained food by pecking 40 times in less than 40 min. Once consistent responding had been achieved, a forced-switching condition was introduced. For the remainder of the experiment, the central key was dark and was not functional.

Forced-Switching Procedure

The purpose of this procedure was to reduce strong side preferences. Dependent concurrent random-interval schedules (RI 60-sec; RI 60-sec) were imposed and implemented as follows. Both side keys were white, except when food was delivered. At the beginning of the session and after each delivery of reinforcement, a probability generator determined which key would be *active* ($p = .5$). Then, every second, the probability generator determined whether the next peck on the active key would be reinforced ($p = 1/30$). Once reinforcement was assigned, the probability generator stopped, and the next peck to the active key was reinforced with 4 sec of access to food. This procedure assigned reinforcement every 30 sec, on average, to one or the other key (averaging one reinforcer per minute for pecks on each key). If a pigeon pecked the nonassigned key exclusively, pecks would be extinguished. Thus, the pigeons were forced to sample both keys (Stubbs & Pliskoff, 1969).

Because reinforcer duration (4 sec of access to food) did not vary regardless of which key was active, this procedure did not expose the pigeon to the PD contingencies; its purpose was simply to reduce strong side preferences.

Sessions ended after 40 reinforcers or 40 min had passed, whichever happened first. This procedure was imposed for 15 sessions, after which the subjects were introduced to the unsignaled condition. All subsequent forced-switching procedures were imposed for 15 sessions for all the pigeons.

Phase 1: Unsignaled and Signaled Conditions

Unsignaled condition (Condition 1). One key was assigned as the cooperate (C) key, the other as the defect (D) key. Both of the keys were illuminated white, except when food was delivered (key color was reserved for the signals to be imposed in the next condition). Figure 2 shows an example of a sequence of C- and D-pecks, the probability of reinforcement for each peck (always .04), and the reinforcer duration that would have been obtained if that peck had been reinforced. Each session started with the computer cooperating (the circled C at the beginning of the illustrative choice sequence of Figure 2). If the pigeon's first peck was on the C-key, the two-response sequence was classified as CC; if the pigeon's first peck were on the D-key, the two-response sequence was classified as CD. After either response, a probability generator determined whether the response would be reinforced ($p = .04$) or not ($q = .96$). On average, one reinforcer was delivered for every 25 responses (random-ratio 25). If the response was reinforced, the amount of reinforcement depended on the last two responses. The downward pointing arrows in Figure 2 indicate the two-peck sequence that each peck completed. This sequence determined the reinforcer duration according to the PD matrix of Figure 1 (reward units equal seconds of hopper access) with the pigeon as Player A and the computer as Player B, playing tit-for-tat.

Figure 3A diagrams the local and global contingencies imposed. The dotted line of Figure 3A shows that with this procedure, the average programmed reinforcer duration increases linearly with percentage of cooperation. Thus, for 100% C-pecks, all reinforcers would be 5 sec; for 100% D-pecks, all reinforcers would be 3 sec; for alternation, half of the reinforcers would be 6 sec and half 2 sec,

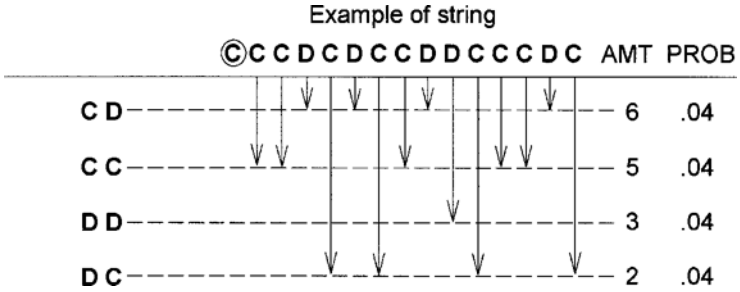

Figure 2. Hypothetical string of C-key and D-key pecks, previous and current choice (CD, CC, DD, or DC), reinforcer duration (AMT), and reinforcement probability (.04 for each peck). Each choice in the string is both the initial and the terminal choices of a two-choice sequence. The circled C indicates that the first choice always counted as a sequence initiated with a cooperation. The arrow pointing downward from each choice indicates the sequence terminated by that choice.

averaging 4 sec. Figure 3B shows average obtained reinforcer duration (seconds per reinforcer) for each session as a function of percentage of cooperation during that session for all 4 pigeons in the IPD conditions over the course of all phases of the experiment. These points are grouped closely around the programmed line (the dotted line of Figure 3A), indicating that the obtained reinforcer amounts approximate the programmed amounts.

Sessions ended after 40 reinforcers or after 40 min had passed, whichever happened first. At the end of the 30th session, the data were examined for stability. If, during the last 5 sessions, the proportions of CC, DD, and changeover (CD plus DC) sequences during each session were within a 0.1 range, the subjects were moved to the next condition; otherwise, the sessions were continued, and stability was checked after every session. Once stability had been attained, a second forced-switching condition and then a signaled condition were introduced.

Signaled condition (Condition 2). This condition was identical to the unsignaled condition, except that after each unreinforced C-peck, *both* keys were illuminated with the cooperation color, which could be green or red (see the third column in Table 2 for the assignment of cooperation colors for each pigeon); after each unreinforced D-peck, *both* keys were illuminated with the assigned defection color. The signals stayed on between successive pecks. As long as the pigeon pecked repeatedly on one key, both keys remained that color. When the pigeon changed over between keys, both keys changed from red to green, or vice-versa. Thus, when signals were present, the color of both keys at any given moment between pecks indicated which key had just been pecked. During reinforcement, both keys were unlit but were illuminated after reinforcement with the color corresponding to the reinforced peck. Session and condition duration were determined using the same criteria as those in the unsignaled condition.

Table 1 indicates the order of experimental conditions and the number of sessions in each condition and phase for each subject.

Table 2 indicates the initial position of the C-key (second column) and the cooperation color during the signaled condition (third column) for each subject.

Phase 2: Test of Control by Signals and Key Location

In this phase, the locations of the cooperation and the defection keys were reversed from those of Phase 1; the C-key during Phase 1 was the D-key during Phase 2, and vice versa. Nevertheless, the relation between the colors of the keys and cooperation/defection did not change. For example, if the red key color (both keys) previously followed a C-key peck (and a choice between probabilistic 5-sec and 6-sec reinforcer durations), it continued to do so after the reversal.

The experimental conditions during this phase were conducted in the following order: a signaled condition (Condition 3) was introduced first, followed by a forced-switching condition and then another signaled condition (Condition 4); in the last condition of this phase (Condition 5), the signals were removed. The stability criterion was the same as that in Phase 1, except in Condition 3, in which the minimum number of sessions was 16 rather than 30, as in the other conditions.[3] (See Table 1 for the number of sessions for each subject in each condition.)

Phase 3: Extensive Exposure to Signaled Contingencies

In this phase (Conditions 6–9), the pigeons were introduced first to a forced-switching condition, followed by a signaled condition in which the C-key and the D-key locations were rereversed to those in Phase 1. In other words, Conditions 6–9 were identical to Condition 2. After each pigeon's choices satisfied the stability criterion at each IPD condition, that condition was terminated, and a forced-switching procedure was introduced. Recall that the forced-switching procedure (concurrent dependent RI 60-sec schedules) embodied neither the IPD contingency (reinforcer duration was always 4 sec) nor the differential signals (both keys were always white). The purpose of the forced-switching procedure was to reduce strong position preferences and expose the pigeons to the contingencies at the beginning of each IPD condition.

Once Condition 6 was finished for Pigeons 2 and 3, they were retired from the experiment. Only the 2 pigeons (1 and 4) that were cooperating least at the end of Condition 4 participated in Conditions 7, 8, and 9.

RESULTS

Phase 1

In the first forced-switching condition, the median percentage of responses on the assigned C-key for the last five sessions ranged from 40% to 57%, for an average across subjects of 46%. When the IPD contingencies were introduced without signals (Condition 1), all the pigeons decreased the proportion of C-key pecks and increased the proportion of D-key pecks. The development of this preference is illustrated in Figure 4, which shows CC and DD peck sequences and changeover responses (equally divided between CD and DC sequences ± 1). The proportion of C-key pecks may be calculated by adding that of CC sequences and half of the changeover

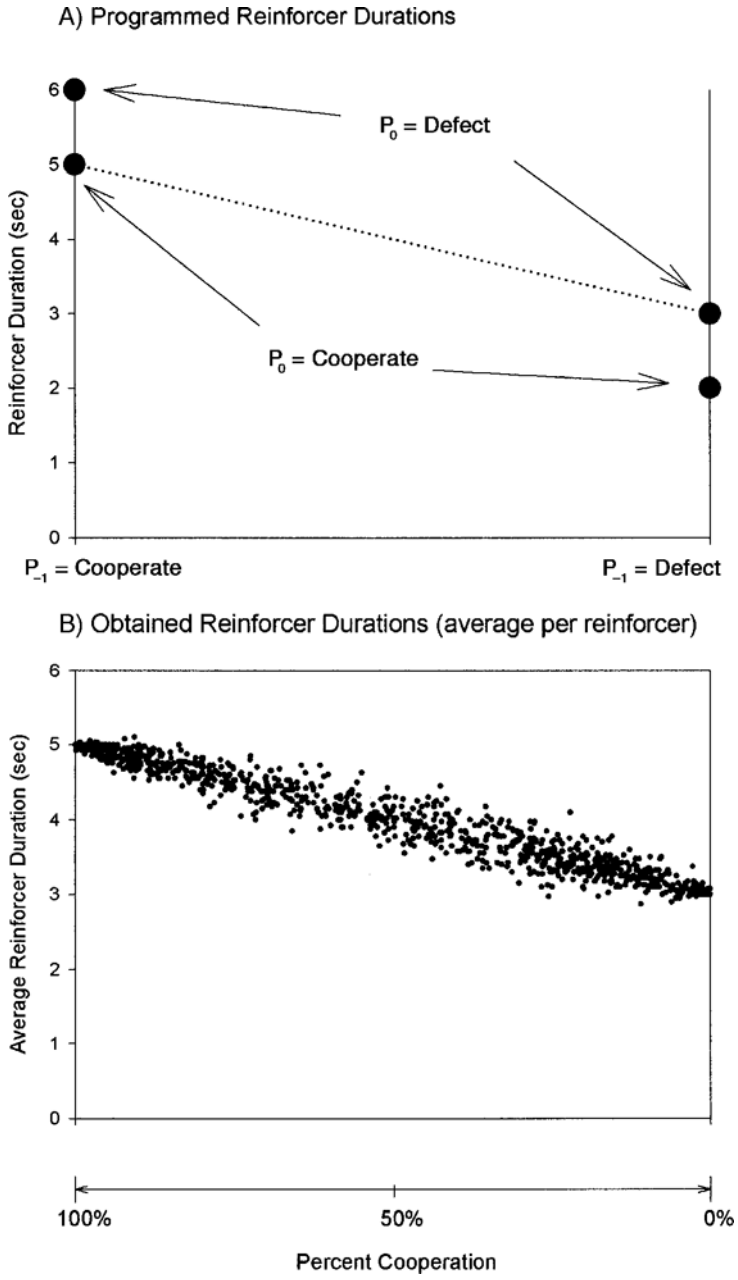

Figure 3. (A) Programmed reinforcer durations and (B) obtained average reinforcer durations (over sessions) as a function of percentage of cooperation.

responses. During the last five sessions of this condition, the median proportion of CC sequences ranged from 0% to 10% (average across subjects = 3%), whereas the median proportion of DD sequences ranged from 75% to 100% (average across subjects = 93%).

During the second forced-switching condition, the median percentage of pecks on the original C-key for the last five sessions ranged from 27% to 43%, for an average across subjects of 37%. After this condition was finished, the signaled IPD contingencies were introduced. As is illustrated in Figure 5, DD sequences initially increased for all the pigeons. For 3 pigeons, this trend was followed by a decline in DD sequences and an increase in CC sequences. Then, during the last sessions of the signaled condition, for 2 of these 3 pigeons, DD sequences rebounded. For each pigeon, the difference between the average proportion of C-key pecks during the last five sessions of the signaled condition (Condition 2) and that

Table 1
Number of Sessions for Each Subject in Each Condition

Condition	C-Key	Subject B 452	B 464	B 391	B 420
Phase 1					
Forced switching	–	15	15	15	15
1. Unsignaled	A	30	30	30	30
Forced switching	–	15	15	15	15
2. Signaled	A	43	76	30	30
Phase 2					
3. Signaled	B	16	61	16	16
Forced switching	–	15	15	15	15
4. Signaled	B	43	35	30	30
5. Unsignaled	B	37	35	31	30
Phase 3					
Forced switching	–	16	15	15	15
6. Signaled	A	30	122	30	30
Forced switching	–	15			15
7. Signaled	A	33			30
Forced switching	–	16			15
8. Signaled	A	30			43
Forced switching	–	15			15
9. Signaled	A	40			30

Note—The second column indicates the side key that functioned as a cooperation (C) key (A = left and B = right for half of the subjects, and vice versa; see Table 2). The order in which the subjects received the experimental conditions goes from top to bottom.

of the unsignaled condition (Condition 1) is presented in the first panel of Figure 6 ("Add Signals").

Phase 2

Figure 7 shows performance after the keys were switched (Condition 3) and after interpolated forced switching (Condition 4). At the end of Phase 1, 2 pigeons (2 and 3) had been mostly cooperating, whereas 2 (Pigeons 1 and 4) had been mostly defecting. The leftmost points of Figure 7 show what happened immediately after the keys were reversed: All the pigeons retained preference for the key that they had been pecking at the end of Phase 1. The pigeons that had been mostly cooperating were now defecting, and vice versa. Clearly, the signals, although they aided in acquisition of cooperation in Phase 1, did not gain control over any pigeon's behavior. Key location, rather than key color, determined initial preference in Condition 3.

Over the course of Conditions 3 and 4, the 2 pigeons (2 and 3) that had been mostly cooperating at the end of Condition 2 and had been mostly defecting at the beginning of Condition 3 began mostly to cooperate again. Of the 2 pigeons (1 and 4) that had been mostly defecting at the end of Condition 2 and had been mostly cooperating at the beginning of Condition 3, 1 (Pigeon 1) began to defect and then to cooperate again during Condition 4 (the cycling accompanied by an elevated level of switching). This pigeon finally reached the stability criterion in Condition 4, with slightly more C-key pecks than D-key pecks. The other pigeon (4) kept cooperating at a high rate through Conditions 3 and 4. Pigeon 4 had evidently developed a strong preference for the defection key in Condition 1 and kept that key preference throughout the experiment to this point. In general, by the end of Condition 4, all 4 pigeons were mostly cooperating; 3 reversed their key preference after the side reversal, whereas 1 (Pigeon 4) maintained a strong preference for what had originally (in Condition 1) been the defection key.

Finally, when signals were removed (Condition 5, Figure 8), 1 pigeon that had been cooperating (Pigeon 3) quickly began to defect; the other pigeon that had been cooperating (Pigeon 4) continued with its strong side preference, whereas 2 others (Pigeons 1 and 2) that had been cooperating kept on cooperating, one (Pigeon 1) actually increasing cooperation and the other (Pigeon 2) decreasing, at the point at which the stability criterion was satisfied.

Table 2
Assignment of Initial C-Key and Color for Each Subject

Subject	Key A	Cooperation Color
Pigeon 1	left	green
Pigeon 2	left	red
Pigeon 3	right	green
Pigeon 4	right	red

Note—When Key A was on the left side, Key B was on the right side, and vice versa. When the cooperation color was green, the defection color was red, and vice versa.

Phase 3

Figure 9 presents the results of Phase 3 (Conditions 6–9). Pigeon 2, whose choices did not attain stability until 122 sessions had been conducted, rapidly reversed key preferences during the initial sessions of Condition 6, but the proportion of C-key pecks declined in subsequent sessions. In contrast, Pigeon 3 persisted in its preference for what was previously the D-key. Both pigeons finished Condition 6 by cooperating in most of the occasions, and no further data were collected.

Pigeons 1 and 4 started Condition 6 with a strong bias toward defection. By Condition 8, both pigeons were producing CC sequences about as frequently as DD sequences. At that stage, Pigeon 1 was switching between the keys at high rates (observed also, only in the same pigeon, during Condition 4). By the end of Condition 9, Pigeons 1 and 4 were producing CC sequences on more than 60% and 90% of their respective choices. The change in the proportion of C-key pecks from the last five sessions of Condition 6 to the last five sessions of Condition 9 is shown in the fourth panel of Figure 6 ("Repeated Forced Switching").

DISCUSSION

Phase 1

Changeover responses, which for 3 of the pigeons had been occurring at a high rate at the beginning of the first unsignaled condition (Condition 1), remained low dur-

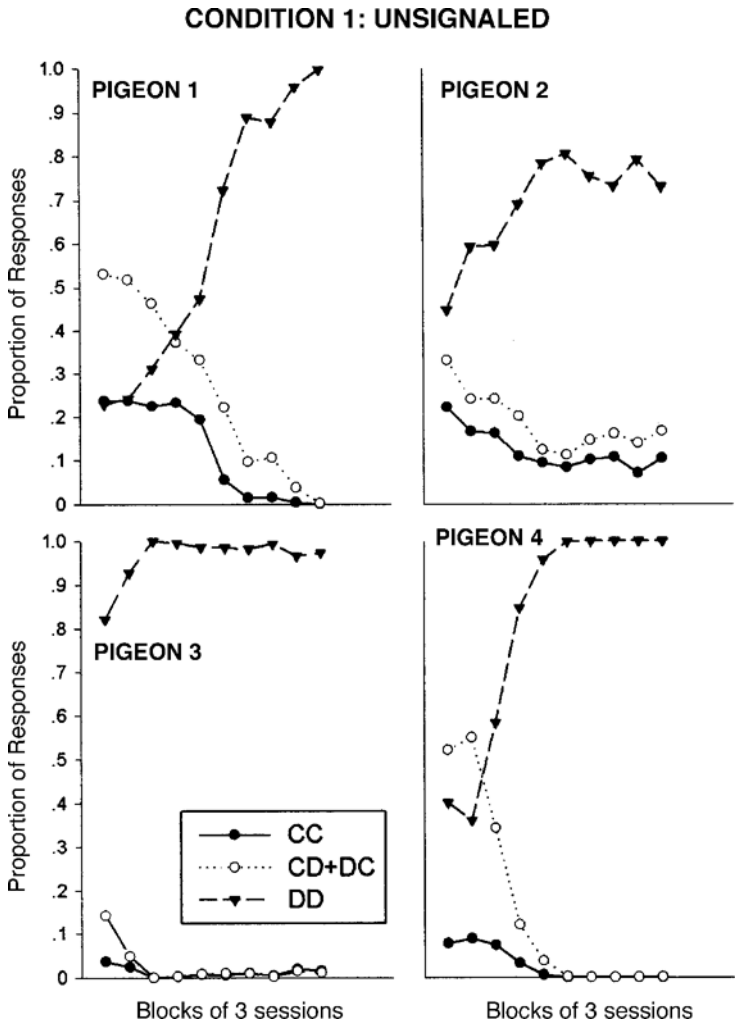

Figure 4. Proportion of CC, DD, and changeover (CD + DC) sequences for Condition 1. Each data point corresponds to the mean proportion of responses in three sessions, except for the first data point, which corresponds, in cases in which the number of sessions was not divisible by three, to data obtained from one session or the mean of two sessions.

ing the following signaled condition (Condition 2). The generally low changeover rate, even without exclusive or near-exclusive preference, indicates that the pigeons emitted long sequences of responses on a given alternative before changing to the other. Although CD sequences were occasionally followed by the highest reinforcer duration (6 sec), DC sequences were occasionally followed by the lowest reinforcer duration (2 sec). This strong difference may have kept the pigeons from switching from strings of D-pecks to strings of C-pecks (and a higher reinforcement rate) and is what creates the dilemma—essentially a self-control dilemma—when playing against tit-for-tat.

Although the consistent defection shown by all 4 pigeons in Condition 1 was disappointing, in a sense, since it indicates that the pigeons were less sensitive to the global than to the local contingencies, defection in Condition 1 also indicates that the cooperation shown by all the pigeons in the later condition was not an artifact of the IPD procedure. For example, in Condition 1, all the pigeons clearly preferred the immediate reward for defection (6 or 3 sec of food access) to that for cooperation (5 or 2 sec of food access); their later choice of the lesser rewards could not have been due to failure to discriminate between 6 and 5 sec or between 3 and 2 sec of food access.

The pattern of choice over time when the signals were first introduced (Figure 5) differs markedly (for 3 of the 4 pigeons) from the pattern without signals. Except for Pigeon 4, signals seem to have caused preference to cycle over sessions, with an initial decline of CC sequences, followed by a rise and then, for Pigeons 1 and

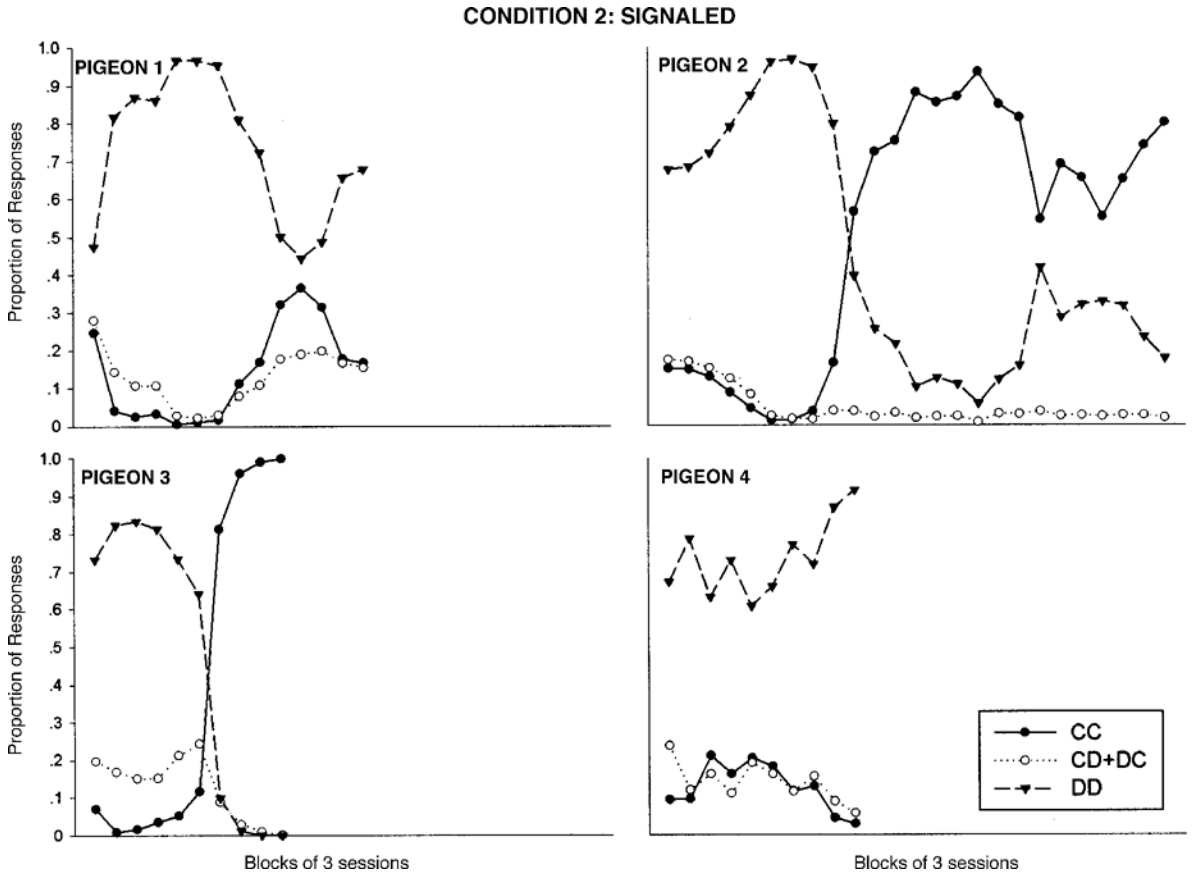

Figure 5. Proportion of CC, DD, and changeover (CD + DC) sequences for Condition 2. Each data point corresponds to the mean proportion of responses in three sessions, except for the first data point, which corresponds, in cases in which the number of sessions was not divisible by three, to data obtained from one session or the mean of two sessions.

2, another drop and then, for Pigeon 2, another rise. The pigeons that cycled the most in preference naturally tended to take longer to reach stability. Since Condition 2 was ended when the stability criterion was satisfied, there is no way to know whether cycling of preference would have been shown by all the pigeons had Condition 2 been further prolonged. In any case, as Figure 6 (first panel) shows, all the pigeons, even Pigeon 4, were cooperating more with signals at the end of Condition 2 than they had been without signals at the end of Condition 1. Two of the pigeons (2 and 3) reversed their initial bias for the C-key and were mostly cooperating by the end of Condition 2. The other 2 pigeons (1 and 4) were still mostly defecting at the end of Condition 2 but had reduced their initial bias toward the defection key.

On the basis of the results of Phase 1, it would be premature to conclude that cooperation and defection were controlled solely by the feedback signals. The pigeons may have acquired a preference for a particular feedback key color (green or red), for pecking on a particular key (left or right), or both in different degrees. To determine the degree of control of key color and key position on the production of C-key peck, the locations of the C-key and the D-key were reversed in Phase 2, and then signals were eliminated.

Phase 2

Although the introduction of signals in Condition 2 had caused all of the pigeons to overcome or reduce their bias toward the defection key, the signals clearly did not gain control over behavior. Even those pigeons (2 and 3) that had learned to cooperate after signals had initially been introduced in Condition 2 and were also cooperating at the end of Condition 4 came to do so gradually, learning to cooperate all over again. However, by the end of Condition 4, all the pigeons were mostly cooperating. Through the reversal and interpolated forced-switching procedure, 3 of the 4 pigeons increased their cooperation. One (Pigeon 4), which was defecting at high levels before the key reversal, actually increased its preference for the initial defection key (now the cooperation key).

Figure 6, second panel, shows the difference between preference for the cooperation key at the end of Condition 4 and the preference for *that key* (formerly the defection key) at the end of Condition 2 (just before the key reversal). Three pigeons increased preference for the

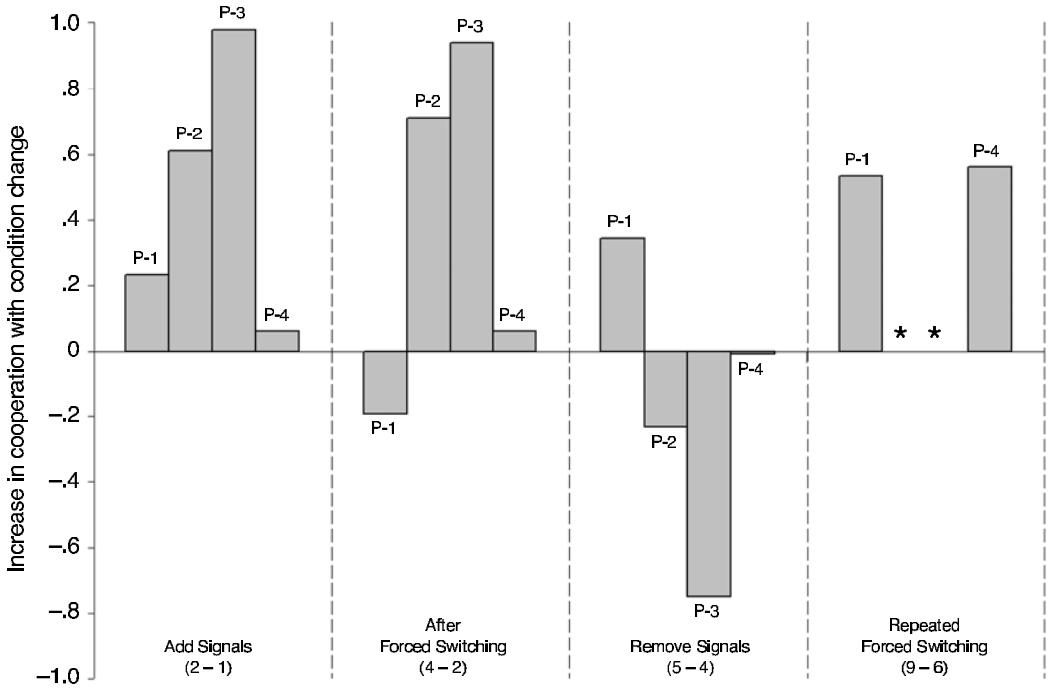

Figure 6. Increases in the proportions of pecks on C-key with condition changes. The numbers in parentheses are condition numbers. The bar heights indicate the proportion of pecks on the C-key in the later condition minus the percentage of pecks on that key in the earlier condition. Asterisks indicate that data were not collected for the corresponding subject in a condition.

new cooperation key, whereas 1 (Pigeon 1) decreased preference for that key. The third panel of Figure 6 shows the effect of removing signals on preference for the new cooperation key. Despite the fact that the signals did not gain control of behavior in Condition 2, 3 of the 4 pigeons decreased preference for the cooperation key, and 1 (Pigeon 1) actually increased preference for the cooperation key.

In summary, the results of Phase 2 show that although the signals did not gain control of behavior, they did promote preference for the cooperation key. More important, after the reversal in sides, 3 of the 4 pigeons eventually reversed preferences along with the contingencies. However, the choices of 2 of the pigeons (1 and 4) were relatively insensitive to the contingencies (note, in Figure 6, Pigeon 4's relative lack of sensitivity and Pigeon 1's unexpected changes in behavior, relative to contingency and signals). Consequently it was decided to pursue extended training with the IPD procedure with these 2 pigeons. In Phase 3, the key positions and signals were restored to their original meanings (those of Condition 2). The IPD and forced-switching contingencies were then alternated repeatedly until, at last, even these relatively contingency-insensitive pigeons learned to cooperate.

Phase 3

The rereversal of signals confirmed the previous results: Choice was not controlled by signals as much as it was by key location. However, with signals in place, all the pigeons came to cooperate on the majority of their choices (it is not known at this point whether they would have done so eventually with the signals absent). Pigeons 2 and 3 attained these levels of cooperation in just one presentation of the signaled condition, although it is unclear whether Pigeon 3 learned to cooperate or whether its performance was an artifact of persistence in key preference. Nevertheless, the proportion of choices on the previous D-key (now the C-key), increased more than 10% in Pigeon 3 from Conditions 5 to 6 (mean of last five sessions of each condition).

The performance of Pigeon 3 in Conditions 3 and 4 had suggested that repeating the signaled condition with an intermediate forced-switching procedure could increase the rate of CC sequences. Data from Conditions 6–9 indicate that such treatment was very effective for Pigeons 1 and 4. It is not fully clear what aspect of this treatment is responsible for the increasing rate of CC sequences, but it seems that when a condition is started with a very high proportion of defections, acquisition of cooperation is retarded, possibly because of lack of contact with the IPD contingencies. The forced-switching procedure tended to pull the proportion of responses a bit closer to indifference at the beginning of the following condition (not apparent from Figure 9, since this change was confined to the very first session after each forced-switching exposure). The forced-switching procedure may thus have increased contact with the IPD

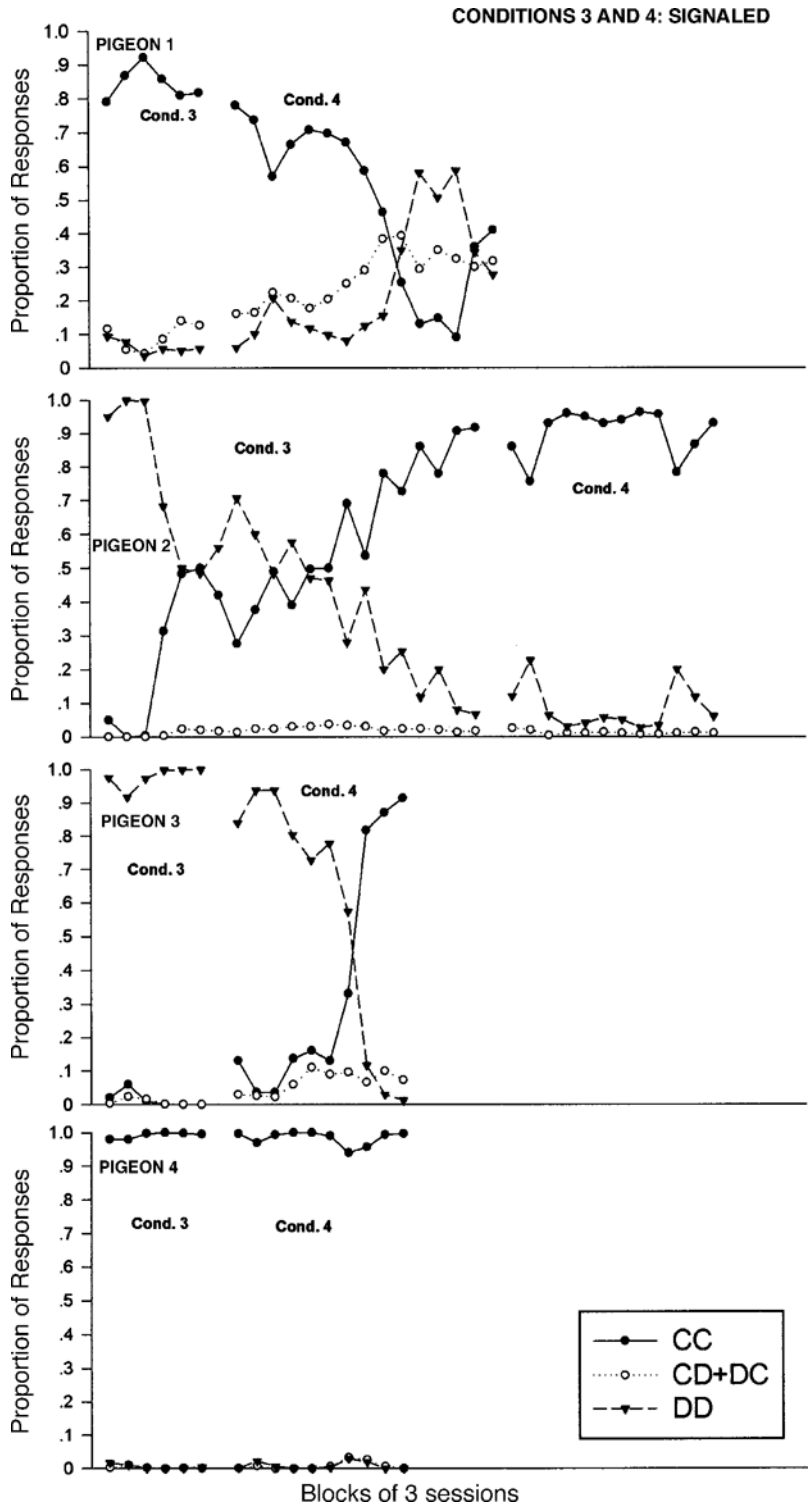

Figure 7. Proportion of CC, DD, and changeover (CD + DC) sequences for Conditions 3 and 4. Each data point corresponds to the mean proportion of responses in three sessions, except for the first data point, which corresponds, in cases in which the number of sessions was not divisible by three, to data obtained from one session or the mean of two sessions.

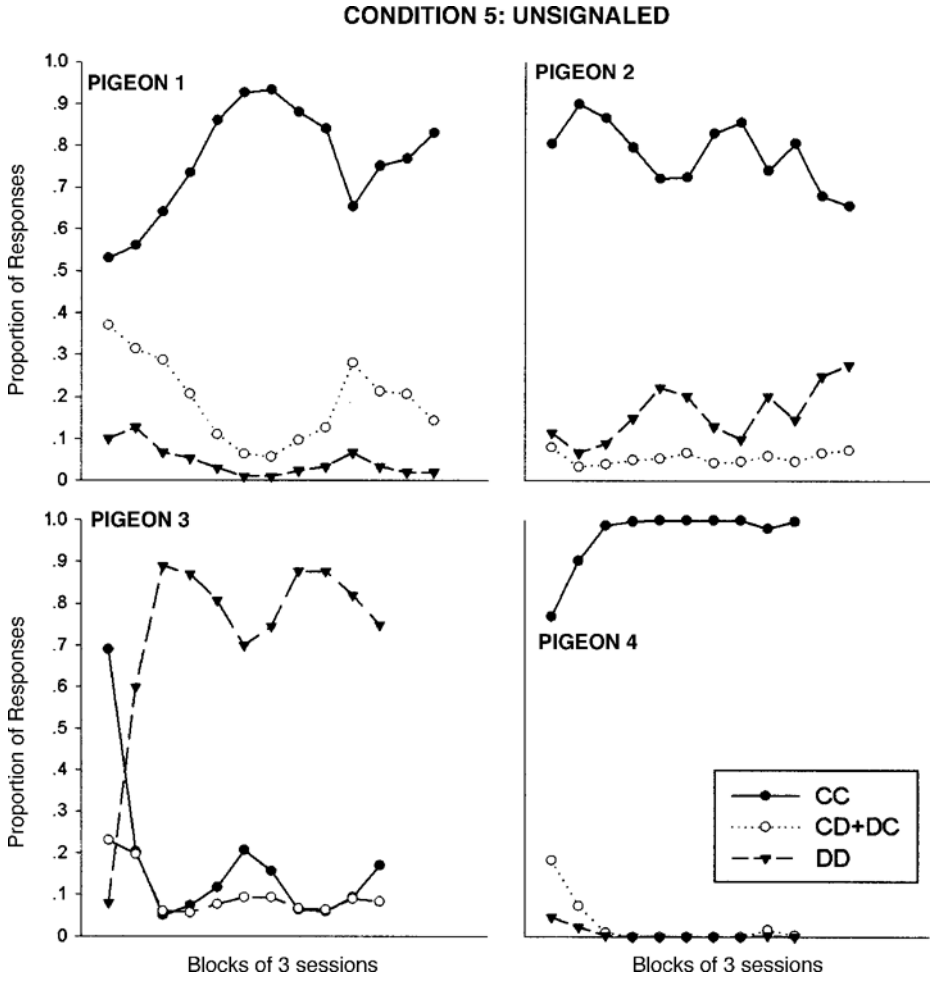

Figure 8. Proportion of CC, DD, and changeover (CD + DC) sequences for Condition 5. Each data point corresponds to the mean proportion of responses in three sessions, except for the first data point, which corresponds, in cases in which the number of sessions was not divisible by three, to data obtained from one session or the mean of two sessions.

contingencies (reinforcement of cooperation and punishment of defection), thereby increasing the rate of CC sequences.

GENERAL DISCUSSION

As the introduction stated, cooperation in an IPD against tit-for-tat is a self-control response under conditions of complex ambivalence (cooperation maximizes global, as opposed to local, reinforcement rate). The consistent defection obtained in the first unsignaled condition (Condition 1) indicates that the pigeons' choices were controlled in that condition by their local, rather than their global, consequences. These results are consistent with Green et al.'s (1995) findings. Since global and local maximizations were determined by the locations of the cooperation and defection keys, the tit-for-tat contingency may be said to have developed a very strong bias for the defection key (left for 2 pigeons, right for the other 2) in Condition 1.

Once the feedback signals were introduced, however, cooperation increased for all the pigeons, reversing the initial bias for 2 pigeons and decreasing it slightly for the other 2. When C-keys and D-keys were reversed in Condition 3, 3 pigeons increased their preference for the C-key (previously the D-key). When the signals were removed in Condition 5, 3 pigeons continued to cooperate on more than half of their choices (1 actually increasing cooperation choices), whereas a 4th defected. Eventually, over the course of several exposures to the forced-switching procedure, even the 2 least contingency-sensitive pigeons came to cooperate on more than 50% of their choices. Thus, with sufficient exposure to the experimental contingencies with feedback signals, all the pigeons learned to cooperate.

Cooperation was more consistent in this free-operant experiment than in Baker and Rachlin's (2002) trial-by-trial experiment (the only other published experiment in which nonhumans learned to cooperate in an IPD against

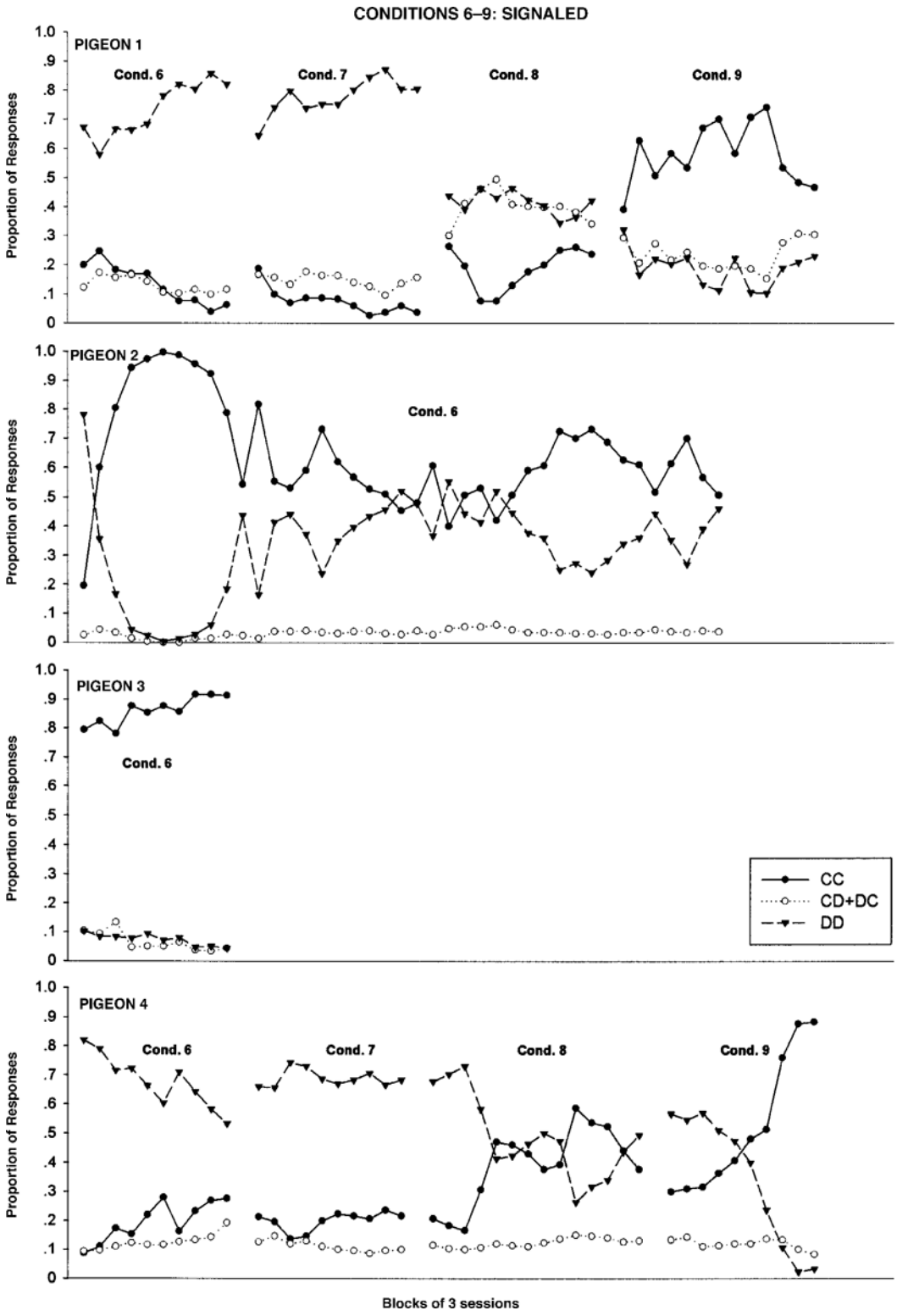

Figure 9. Proportion of CC, DD, and changeover (CD + DC) sequences for Pigeons 1–4, Conditions 6–9. Each data point corresponds to the mean proportion of responses in three sessions, except for the first data point, which corresponds, in cases in which the number of sessions was not divisible by three, to data obtained from one session or the mean of two sessions.

tit-for-tat).[4] One explanation for the effectiveness of the operant procedure would rely on the many choice sequences without intervening reinforcement; signals corresponding to the choice just previous to the reinforced choice may thus have been more salient than they would have been in a trial-by-trial procedure. Since the pair of larger alternatives (6 or 5 sec of food access) were obtained by cooperating on the just previous choice, cooperation frequency may have increased as the saliency of the signals increased.

Another explanation for the effectiveness of the operant procedure in producing cooperation relies on the conception of probability as equivalent, in certain respects, to delay (Rachlin, Logue, Gibbon, & Frankel, 1986). For a random-ratio schedule with reinforcement probability, p, the average delay (or waiting time) between a keypeck and a reinforcement is $t[(1/p)-1]$, where t is the average interresponse time. With the IPD matrix used here, defection was reinforced with an increase of 1 sec of food access (6−5 or 3−2), but cooperation on the just prior choice was reinforced with an increase of 3 sec of food access (6−3 or 5−2). The effective choice in the present experiment was, therefore, between a smaller–sooner reward increase (1-sec difference delayed by $t[(1/p)-1]$ seconds) and a larger–later reward increase (3-sec difference delayed by $t[(1/p)-1] + t$ seconds). In the present experiment, $t \cong 0.5$ sec and $p = 1/25$. Thus, the average delay between a response and (its own) reinforcement was 12 sec, whereas the average delay between the just prior response and reinforcement (of the subsequent response) was 12.5 sec. Therefore, the operant procedure (i.e., the random-ratio schedule) added 12 sec to both smaller–sooner and larger–later alternatives, as in a commitment procedure (Rachlin & Green, 1972), in which such addition sharply increased choice of the larger–later alternative. An equivalent procedure with probabilistic reinforcers (and human choosers) increased choice of a larger, less probable alternative (Rachlin, Castrogiovanni, & Cross, 1987). This translates, in the present experiment, into increased cooperation.

This second alternative explanation for increased cooperation with the operant procedure does not account for the role of the signals, which were demonstrably important in this experiment. It may be that both explanations are correct to some degree—the former accounting for the learning of cooperation, the latter for its maintenance. This would explain why signals were evidently necessary for the learning of cooperation but did not gain control over behavior.

The results should be interpreted in the light of the fact that with very clear and detailed signals of the contingencies, including a visual representation of the entire matrix in Figure 1, 10%–20% of undergraduate subjects failed to learn to cooperate versus tit-for-tat over as many as 100 trials (Baker & Rachlin, 2001; Brown & Rachlin, 1999). It may be claimed that the pigeons in the present experiment had many more than 100 trials and so might have been expected to do better than humans. But the pigeons did not have a long history of social interaction, an extensive verbal repertory, or experience playing games against a computer, as all of our human subjects had.

The tit-for-tat contingency is a much more difficult self-control problem than simple choice between a smaller–sooner and a larger–later reward. Cooperation versus tit-for-tat requires the choice of, the obtaining of, and (in the case of nonhuman subjects) the consumption of a smaller, as opposed to a larger, immediate reward before the long-term benefit is realized. This requirement makes the self-control problem one of local versus global contingencies, rather than immediate versus delayed contingencies. But as Heyman and Tanz (1995) and Baker and Rachlin (2002) have shown, when global contingencies are clearly signaled, even the behavior of pigeons may come to be controlled by those contingencies.

The introduction of the signals in this experiment made the problem much easier than it was without signals. A given key color always indicated whether a pigeon had just chosen to cooperate or to defect—hence, whether the pair of larger alternatives or the pair of smaller alternatives was in effect. These contemporaneous signals may have substituted, with nonhumans, for the extensive verbal repertoire that humans use to solve corresponding self-control problems. However, the precise role of signals in this experiment is still unclear. On one hand, all the pigeons defected almost exclusively until signals were introduced, and when signals were removed, 3 of the 4 pigeons reduced their rates of cooperation. On the other hand, when signals were reversed, none of the pigeons immediately reversed its key preferences. The signals seem to have served as an aid in reversing preference from the defection key to the cooperation key (thereby increasing overall reinforcement rate), but it was the key location, rather than key color, that immediately controlled choice.

The unsignaled conditions (1 and 5) were not imposed as frequently as the signaled conditions for any of the subjects. We therefore cannot be certain that, had we persisted, even without signals, cooperation would not have eventually been acquired. It is not clear, therefore, how much of the success in inducing cooperation in the two least contingency-sensitive pigeons in Phase 3 was due to the interaction between signals and repeated exposure to the contingencies and how much was due to the repeated exposure to the contingencies themselves. These are matters to be sorted out in further experimentation.

REFERENCES

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

Baker, F., & Rachlin, H. (2001). Probability of reciprocation in repeated prisoner's dilemma game. *Journal of Behavioral Decision Making*, **14**, 51-67.

Baker, F., & Rachlin, H. (2002). Self-control by pigeons in the prisoner's dilemma. *Psychonomic Bulletin & Review*, **9**, 482-488.

Brown, J., & Rachlin, H. (1999). Self-control and social cooperation. *Behavioural Processes*, **47**, 65-72.

Green, L., Fristoe, N., & Myerson, J. (1994). Temporal discounting and preference reversals in choice between delayed outcomes. *Psychonomic Bulletin & Review*, **1**, 383-389.

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.

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.

Herrnstein, R. J., & Prelec, D. (1992). A theory of addiction. In G. Loewenstein & J. Elster (Eds.), *Choice over time* (pp. 331-360). New York: Russell Sage Foundation.

Heyman, G. M., & Tanz, L. (1995). How to teach a pigeon to maximize overall reinforcement rate. *Journal of the Experimental Analysis of Behavior*, **64**, 277-297.

Komorita, S. S. (1994). *Social dilemmas.* Dubuque, IA: Brown Communications.

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 analysis of behavior: Vol. 5. The effect of delay and 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* (Vol. 40, pp. 327-364). San Diego: Academic Press.

Rachlin, H., Castrogiovanni, A., & Cross, D. (1987). Probability and delay in commitment. *Journal of the Experimental Analysis of Behavior*, **48**, 347-353.

Rachlin, H., & Green, L. (1972). Commitment, choice and self-control. *Journal of the Experimental Analysis of Behavior*, **17**, 15-22.

Rachlin, H., Logue, A. W., Gibbon, J., & Frankel, M. (1986). Cognition and behavior in studies of choice. *Psychological Review*, **93**, 33-45.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation.* Ann Arbor: University of Michigan Press.

Reboreda, J. C., & Kacelnik, A. (1993). The role of autoshaping in cooperative two-player games between starlings. *Journal of the Experimental Analysis of Behavior*, **60**, 67-83.

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., McLinn, C. M., & Stevens, J. R. (2002). Discounting and reciprocity in an Iterated Prisoner's Dilemma. *Science*, **298**, 2216-2218.

Stubbs, D. A., & Pliskoff, S. S. (1969). Concurrent responding with fixed relative rate of reinforcement. *Journal of the Experimental Analysis of Behavior*, **12**, 887-895.

NOTES

1. A constraint that must be imposed on the PD payoff matrix for cooperation to be the highest long-term outcome in tit-for-tat is that the "good" outcome (5 reward units in Figure 1) must be preferred over the average of the "best" and the "worst" outcome (4 reward units in Figure 1). Otherwise, Player A in the example could maximize the long-term outcome by alternately cooperating and defecting.

2. The increased amount of food obtained for a cooperation was delivered only after the following trial (26 sec) and the delay between choice and food delivery (6 sec) had elapsed.

3. Stability was checked earlier in this condition because of what it was designed to test. The degree of control of the feedback signals on the already acquired preference for CC sequences was expected to be determined after fewer sessions than was the acquisition of this preference.

4. The procedure followed by Stephens et al. (2002), in which cooperation was maintained with blue jays, also delayed reinforcement, but the delays were equal for cooperation and defection (all reinforcers accumulated over four choices). Unless the dropping of pellets into a transparent container is considered to be an effective reinforcer, the procedure provided no incentive to defect.

(Manuscript received April 28, 2003;
revision accepted for publication July 23, 2003.)