Internal Self-Control in Pigeons[1]

George Ainslie

Harvard University[2]

April, 1982

ABSTRACT

Seven pigeons received 2 seconds' access to grain if they pecked a lighted key, but 4 seconds' access to grain if they withheld pecks for X seconds (a parameter) until the key went dark. Over four years of daily trials with this differential reinforcement for not responding (DRO) schedule, only three birds reduced their tendency to peck the key; this change occurred suddenly, after varying numbers of sessions, and it proceeded rapidly until a markedly lower, subsequently stable level of pecking was established. This decreased tendency to peck represents a partial correction of the undervaluation of delayed events described by Herrnstein's matching law; it was not accompanied by observable signs of behavioral pre-commitment, and might have occurred either through wholly covert precommitment or an actual change in the future discount function.

Herrnstein's matching law (1970) predicts that the effectiveness of alternative rewards that are available at different times will be distorted relative to their objective values as the earlier reward becomes imminently available (Ainslie, 1975). Such a distortion could be expected to produce problems of adaptation in many circumstances; for instance, a pigeon might markedly reduce its food intake by choosing smaller-earlier over larger-later feedings. In fact, pigeons are known to prefer food reinforcers that are available immediately over substantially larger food reinforcers at only a few seconds' delay (Ainslie and Herrnstein, 1981; Rachlin and Green, 1972). However, the matching law suggests two ways in which compensation for this distortion might occur: (1) Organisms might learn to nullify the temporary dominance of poorer but imminent rewards by precommitting tactics, like Ulysses tying himself to the mast to get by the Sirens (Ainslie, 1974, 1975, and in press); or (2) Organisms might learn to vary the slope of the discount curve itself as an operant response, becoming more responsive to the objective contingencies of reward in some or all situations.

Pigeons have been shown to choose physical precommitting devices when such devices have been made available in advance (Ainslie, 1974; Ainslie and Herrnstein: 1981; Rachlin and Green, 1972). Moreover, Mazur and Logue have shown that pigeons may come to choose a larger, later reinforcer despite the continuing availability of an immediate, smaller alternative (1978). They shaped this preference by starting with small differential delay requirements and gradually increasing them: two of their four experimental subjects came to choose 6 seconds' access to grain delayed 6 seconds over an immediate 2 second access to grain and retained this preference when the keys producing each outcome were reversed. The author likewise has observed occasional quantum jumps in individual subjects' tendency to wait for larger reinforcers during differential reinforcement for not responding (DRO); such jumps have occurred in a minority of subjects and after widely varying numbers of trials.

Increases in gratification delay without the presence of physical precommitment might seem to be evidence for learned changes in the discount function itself, but they could also occur because subjects had learned a non-physical precommitting device. The clinical literature on human impulse control contains three kinds of intrapsychic precommitting tactics, none of which are implausible in pigeons (Ainslie, in press):

1. Subjects may divert their attention away from information that will tell them when the smaller, earlier reward is available. Pigeons using this tactic might turn away from the key so as not to see it, although they might not have to change their behavior in any observable way.

2. Subjects may develop a behavior incompatible with key-pecking, which might be observable: pecking the chamber wall, grooming etc.

3. Subjects might begin to use their choice on each trial as a predictor of their probable reinforcements on subsequent trials, thus bringing the differential reinforcing effect of the whole series of alternative reinforcers to bear on each choice (Ainslie, 1975). This perceptual process would be an analog of private rule-making in people, e.g., where a person stakes the credibility of his new diet against each temptation to violate it. If this process occurs in pigeons, it should produce a positive feedback effect: Gratification deferring choices should increase the expectable aggregate reward at stake in future choices, thereby increasing the likelihood of gratification-deferring choices on subsequent trials; conversely, immediate consumption choices should decrease aggregate expected reinforcement at stake and thus decrease the likelihood of gratification-deferring choices on subsequent trials. This positive feedback effect should manifest itself in a tendency for choices in each direction to occur in runs, that is, for there to be a higher conditional probability of a behavior's occurrence on the subsequent trial if it has occurred on the current one.

The present research was designed to study spontaneous increases in pigeons'

gratification-deferral behavior. Important questions about such increases include: 1) Will they occur in all subjects? 2) Does shaping contribute to their occurrence? 3) How much can gratification deferral be increased? 4) What is the usual time course of learning? and 5) Are the increases stable? While deferral behavior was being charted, information was also gathered to discriminate precommitment from a modified discount function per se as the learned deferral mechanism.

Pigeons were required to withhold pecks to obtain larger, later reinforcers. The subjects obtained most of their daily food from this source over a period of more than 4 years, a large part of their life spans, while changes in the incidence of their pecks were recorded.

EXPERIMENT 1

METHOD

Subjects. Seven white Carneaux pigeons with no prior experience in differential reinforcement for not responding (DRO), differential reinforcement of low rates (DRL), or other response-suppressing contingencies were kept at 80% of their free feeding weight.

Apparatus. The subjects performed in a sound-proof chamber that measured 30 by 32 by 33 cm. In one wall was a single key, which was centered at a height of 24 cm and required 14 g (1.4 N) to operate. The key could be illuminated from behind by red 7-W Christmas-tree bulbs. Ten cm beneath the key was a 5 by 6 cm niche in the wall, into which a hopper of grain could be raised by an electromagnet. Centered in the top of the chamber was a one-way viewing lens, of the kind often set into the front door of apartments. It measured 1 cm in diameter and permitted the experimenter to see most of the chamber without being seen from inside. No houselight was used. White noise was piped into the box to mask environmental sounds.

Procedure. Trials lasted 20 seconds each regardless of subjects' behavior. In

daily sessions of 40 continuous trials each, the key was transilluminated for X seconds (a parameter). If a subject pecked a key while it was lit, he was immediately given two seconds' access to grain. If he did not peck, he received four seconds' access to grain as soon as the key went dark. The incidence of pecks was measured at each value of X.

Design. Subjects were divided into two groups. For Group A (3 birds) the value of X remained at three seconds over 53 blocks of 5 sessions each. For Group B (4 birds), X was set at three seconds for 3 blocks, then changed to one second and varied so as to maintain a 15-30% incidence of pecks: When a subject averaged pecking on more than 30% of trials over a 5 session block (more than 12 out of 40 trials per day), the value of X was decreased by .2 seconds for the next block; if a subject averaged fewer than 15% peck trials in a block (fewer than 6 pecks per session), the value of X was raised by .2 seconds for the next block. If X had not risen to 3 seconds after 47 blocks of sessions, X was set at 3 seconds for 3 blocks in order to compare the performance of Group B with Group A.

If a subject reached low levels of pecking at X = 3 seconds, it was challenged with a procedure to test the stability of this new rate of behavior: After pecking on fewer than 30% of trials at X = 3 seconds for 3 blocks in a row, it was given blocks of sessions with X = 10 seconds to bring its pecking incidence back to 100%, alternating with test blocks with X = 3 seconds.

Sample observations were made of subjects' behavior in the box, to see if subjects were turning away from the key or engaging in behaviors incompatible with pecking.

RESULTS

Within a few sessions from the start, all subjects' incidence of pecking had risen to over 80% at X = 3 seconds. However, all subjects continued to withhold

pecks on some trials, and were thus repeatedly exposed to the larger, later reinforcement contingency. Behavior for the next year in about 10,000 trials is shown in Figure 1. The majority of subjects in both the constant delay group and the shaping group developed little or no tendency to withhold pecks. One bird in each group (no. 3 and no. 5) suddenly developed a much greater tendency to withhold pecks, which developed over a few weeks after a long period of relatively level pecking rates.

After the two birds which reduced their tendency to peck had been challenged with 10 second delay periods, they almost always returned quickly to low rates of pecking at X = 3 seconds (Figure 1c). However, both showed some signs of instability. Bird 3 once continued to peck on 100% of trials after X was returned to 3 seconds; a block of trials where X = 1.5 seconds produced exposure to the larger, later reinforcer and was followed by low levels of pecking when X again = 3 seconds and after subsequent challenges where X = 10 seconds. Bird 5 began pecking more often after 3 challenges, and never returned to its previous rate of less than 30%, although it never returned to its original rate of more than 80%.

Subjects rarely turned away from the key. They often pecked at the chamber wall near the key before or instead of pecking the key, but this occurred both in subjects which increased their gratification-deferral and those which did not. The gross appearance of this behavior did not change when subjects increased their deferral.

EXPERIMENT 2

METHOD

Subjects. After the completion of Experiment 1, the 7 subjects were begun on a similar experiment, modified so that the tendency of peck trials (or non-peck trials) to occur in runs could be observed.

Apparatus. The apparatus described in Experiment 1 continued in use.

Procedure. The contingencies of reinforcement were similar to those described in Experiment 1. However, the length of each trial was fixed at 30 seconds. The delay period, X, was varied stepwise among 1 second, 1.5 seconds, 2 seconds, 3 seconds, 4 seconds, 5 seconds, 6 seconds, 8 seconds, 10 seconds, and 12 seconds. Part way through the experiment the lower end of the range was extended to 0.7 seconds. The value of X was usually changed after each block of 5 sessions; however, it was sometimes held at the same value for several blocks of sessions to collect more observations at particular rates of pecking, or when pecking had not reached a stable rate. The direction of change of X was not reversed until a subject's behavior was at least 85% in one direction for a block of five sessions.

Two control conditions were designed in such a way that subjects received either 4 seconds' access to grain or nothing, so that there were no smaller, earlier reinforcers that might have been temporarily preferred and thus no reason to expect peck trials (or non-peck trials) to occur in runs. After one iteration of X from low to high (or high to low) values and back (about 150 sessions), the procedure was changed to a differential reinforcement of low rates (DRL) contingency for 50 sessions at a value that elicited moderate pecking; subjects then returned to the previous DRO contingency for about two more iterations. Finally, subjects faced a DRL 2 second (birds 1 and 2) or 3 second (birds 3 - 5, 7) schedule, followed by the DRO schedule of the same time value, followed by a schedule in which subjects had to peck so rapidly within the same time period that the success rate was comparable to that on the DRL and DRO schedules with that time value ("Rapid Peck"). Bird 6 died during the third iteration.

Again, samples of subjects' gross behavior were observed. In addition, records were kept of subjects' choices on each trial, so that conditional probability of choices as a function of previous choices could be calculated.

RESULTS

Figure 2 shows the effect of varying X over 2½ to 3½ cycles per subject (DRL data are not comparable and are omitted). The most striking effect is hysteresis (momentum): In 5 birds (3-7), the pecking rate was much higher at a given value of X if the preceding value was one that led to high pecking rates than if it was one that led to low pecking rates. Birds 1 and 2 may have shown this effect weakly, but their ascending and descending curves were almost superimposed. No bird showed the opposite phenomenon, behavioral contrast, under these circumstances.

When values obtained while X was moving in the same directions are compared from cycle to cycle, the pecking rates are stable for most birds. Over the 3 years that this experiment was run, only 1 subject showed an obvious change in its tendency to withhold pecks: After 7000 trials bird 7, which had shown no decrease in pecking in Experiment 1, made a quantum jump from 50% pecking at 2-4 seconds to 50% pecking at 6-9 seconds, a pattern which remained stable for the rest of the experiment.

As in Experiment 1, no obvious waiting behavior distinguished delayers from non-delayers. Data on the occurrence of runs was essentially negative. Although subjects had a significant tendency to respond in runs in the DRO condition and not the DRL condition, this tendency did not differ from that in the rapid peck condition. Furthermore, it was not clearly greater in the birds that withheld pecks more nor did it increase in bird 7 when peck-withholding increased (Table 1).

DISCUSSION

As methods of quantifying preference, both two-key choice and one-key DRO procedures have major drawbacks: With the discrete trial designs necessary to study actual delays, two-key choice schedules lead to stubborn side preferences; single key DRO schedules add the unknown reinforcing effect of the opportunity to

peck a key in the presence of food to the smaller, earlier food reinforcers themselves. Thus both procedures are prone to underestimate subjects' capacities to learn increased gratification delay. Nevertheless, data from both Mazur's and Logue's two-key experiment (1978) and the present one-key experiment suggest that there is some plasticity in pigeon's tendency to discount future reinforcers. Some pigeons increased their tendency to wait for larger, later reinforcers over periods of months to years. Although neither experiment used a group to control for simple aging, Logue and Mazur re-tested their subjects after a year and found no further change in pecking (Reference Note 1); it is likely that the experimental history was responsible for the changes reported here.

The research reported here does not show that shaping made any difference in the incidence or speed of these changes; but the number of subjects is too small to completely rule out shaping as a factor. All that can be said with certainty is that in this DRO procedure shaping was not necessary for a great change in peck withholding to occur, and in some subjects it was not sufficient, even over a four-year experimental history. However, Mazur and Logue's experiment suggests that at least some kind of challenge with a gratification delay problem is necessary for this change to occur (1978). When their control subjects spent time choosing between 6 seconds access to grain at 6 seconds delay and 2 seconds access to grain at 5.5 seconds delay, they regularly chose the former option and subsequently failed to delay gratification when 2 seconds of grain immediately was offered as an alternative to 6 seconds of grain at 6 seconds of delay.

When a change in pecking incidence was observed, it had three important features: (1) The change was abrupt, occurring after a plateau of responding that was stable except for hysteresis, and leading to another plateau of responding which was fairly stable but might regress toward the original level (as in bird 5);

Mazur and Logue's two successful subjects also seemed to find a sudden "solution," although the procedure made it hard to observe the time course of learning. Both showed an initial decrease in frequency of gratification deferral as the problem was made harder, followed by a steady increase in deferral as the problem continued to be made harder (the inflection point occurring at 6 seconds of grain after 6 seconds delay vs. 2 seconds of grain at 1.25 seconds delay for their subject 46, and at 6 seconds of grain at 6 seconds delay vs. 2 seconds of grain at 1.75 seconds delay for their subject 291).

(2) For any individual subject one reduction in pecking exhausted the possibility for change; that is, when a subject had made a quantum jump to a lower frequency of pecking, it did not make later jumps to still lower frequencies after another plateau of stable responding.

(3) Subjects which developed low pecking rates showed much more hysteresis than the subjects which did not. That is, for these subjects, recent experience with a schedule that induced a high pecking rate partially abolished the new, low pecking rates at a given delay.

There is now abundant evidence that organisms discount the future in a steep, highly bowed curve (deVilliers and Herrnstein, 1976; Navarick and Fantino, 1976); that they choose precommitting devices that compensate for this steep discounting rate if offered them in advance (Ainslie, 1974; Ainslie and Herrnstein, 1981; Rachlin and Green, 1972); and that they sometimes compensate for it even when no precommitting device is offered by the experimenter (Mazur and Logue, 1978 and the present research). As has been mentioned, this "internal" compensation might occur in two ways: The organism might learn to directly make its discount curve more shallow, or it might find devices to forestall in advance its predictable, temporary preference for smaller, earlier alternatives.

To learn to discount the future in a more shallow curve, an organism would have

to modify the central process that determined the rate of discounting, such as the one described by factor I in Herrnstein's equation (1981):

$$B_1 = \frac{kR_1A_1}{R_1A_1+r_e+D_1I}$$

(where B_1 is the rate of emitting behavior 1, k is the asymptote of B_1, R_1, A_1 and D_1 are the rate, amount, and delay of reinforcement for behavior 1, r_e is the rate of reinforcement for not behavior 1, and I is an empirical constant). This hypothesis has the disadvantage that it requires the organism not to have learned until the DRO experiment a behavior for which a strong differential reinforcement has always existed: Reducing the rate of discounting a future reinforcer makes that reinforcer more effective in the present. Any behavior which had this effect ought to be reinforced by this increase in the current effectiveness of the reinforcer in question. If such behavior could be learned, it should have been shaped to a maximal rate of emission by an experimental subject's ordinary experience before the DRO experiment. The possibility remains, however, that it is an obscure response, which subjects could not find by trial and error until faced with the particular contingencies of this experiment.

To precommit its behavior in advance, an organism would have to find one of the three kinds of device described in the introduction at a time when its behavior was still under the control of the later, larger reinforcer. Many of these devices should entail observable behaviors. It was hoped that the research described here would be a critical experiment which could demonstrate the learning of some precommitting device that was not supplied by the experimenter. It has failed to do so, although the occurrence of purely intrapsychic precommitting behaviors is not entirely testable with available methods.

Even if the delaying behavior is entirely intrapsychic, that is, if no behavior is observable outside of the increased choice of later, larger alternatives, it

should still be feasible to discriminate empirically between the two possible mechanisms. Once learned, a shallower discount curve should be stable ever afterwards, since it will increase the effectiveness of, and thus be reinforced by, all reinforcers that are not imminently available. Precommitment, on the other hand, should be subject to marked disinhibition effects when the reinforcers or circumstances of reinforcement are changed. Furthermore, precommitment must be performed in advance, and should be undermined by schedules that make smaller-earlier reinforcers available without warning or available continuously during each session.

The increased gratification deferral observed here was relatively stable, but only in Experiment I was it challenged with contingencies likely to produce disinhibition. There, alternation between 3 and 10 second delay requirements disrupted gratification deferral at least temporarily in both birds that faced it, but this was a partial effect, qualitatively different in each bird, and clearly nothing that could support a major conclusion.

At present, the question of how pigeons increase their ability to deter gratification remains open.

FOOTNOTES

[1]Warm thanks to Richard Herrnstein, who made his laboratory resources freely available to the author; to John Cerella who helped write the control programs; and to Virginia Upham, who did most of the bird running.

[2]Now at the Life Management Study, VA Medical Center #151, Coatesville, PA 19320.

REFERENCES

Ainslie, G. Impulse Control in Pigeons. Journal of the Experimental Analysis of Behavior 21:485-489, 1974

Ainslie, G. Specious Reward: A Behavioral Theory of Impulsiveness and Control. Psychological Bulletin 82:463-496, 1975

Ainslie, G. A Behavioral Economic Approach to the Defense Mechanisms: Freud's Energy Theory Revisited. In: Elster, J. (Ed.) The Multiple Self. Cambridge, U.K.: Cambridge University, in press

Ainslie, G. & Herrnstein, R. Preference reversal and delayed reinforcement. Animal Learning and Behavior 9:476-482, 1981

deVilliers, P. & Herrnstein, R. Toward a law of response strength. Psychological Bulletin 83:1131-1158, 1976

Herrnstein, R. On the law of effect. Journal of the Experimental Analysis of Behavior 13:243-266, 1970

Herrnstein, R. Self-control as response strength. In: Bradshaw, C., Szabadi, E. & Lowe, C., (Eds.) Quantification of Steady State Operant Behavior. New York: Elsevier, 1981

Logue, A. W. & Mazur, J. E. Unpublished manuscript, 1979

Mazur, J. E. & Logue, A. W. Choice in a "self-control" paradigm: Effects of a fading procedure. Journal of the Experimental Analysis of Behavior 30:11-17, 1978

Navarick, D. J. & Fantino, E. Self-control and general models of choice. Journal of Experimental Psychology: Animal Behavior Processes, 1976, 2, 75-87

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

TABLE 1

Percent of sessions in which $\frac{\#\text{ of NO-PECK trials following a NO-PECK trial}}{\#\text{ of PECK trials following a NO-PECK trial}} > \frac{\#\text{ of NO-PECK trials following a PECK trial}}{\#\text{ of PECK trials following a PECK trial}}$

(This measure of tendency to respond in runs is independent of the absolute number of peck and no-peck trials, but trials which did not contain at least one of each sequence were omitted. Number of trials included is in parentheses).

Bird	1	2	3	4	5	6	7
DRO (total)	62 (273)	76 (319)	77 (203)	73 (211)	53 (319)	80 (174)	77 (201)
DRL (total)	42 (74)	42 (79)	31 (55)	39 (80)	57 (61)	76 (42)	41 (29)
DRL (last run)	33 (48)	35 (40)	29 (31)	53 (34)	37 (38)		56 (18)
DRO (last run)	50 (16)	64 (33)	73 (33)	86 (7)	58 (19)		64 (36)
Rapid peck	100 (9)	70 (10)	48 (25)	78 (11)	79 (19)		75 (12)

FIGURE CAPTIONS

1. a. Pecking rates of subjects with constant 3" key duration.

   b. Key durations which kept pecking rate at 15-30% of trials: percent of trials with a peck when key duration returned to 3" in parentheses.

   c. Pecking rates of the two subjects which had developed low pecking rates at 3" key duration, when key duration alternated between 3" and 10".

2. Pecking rates as a function of key duration; X denotes the start of an iteration.

Figure 1A

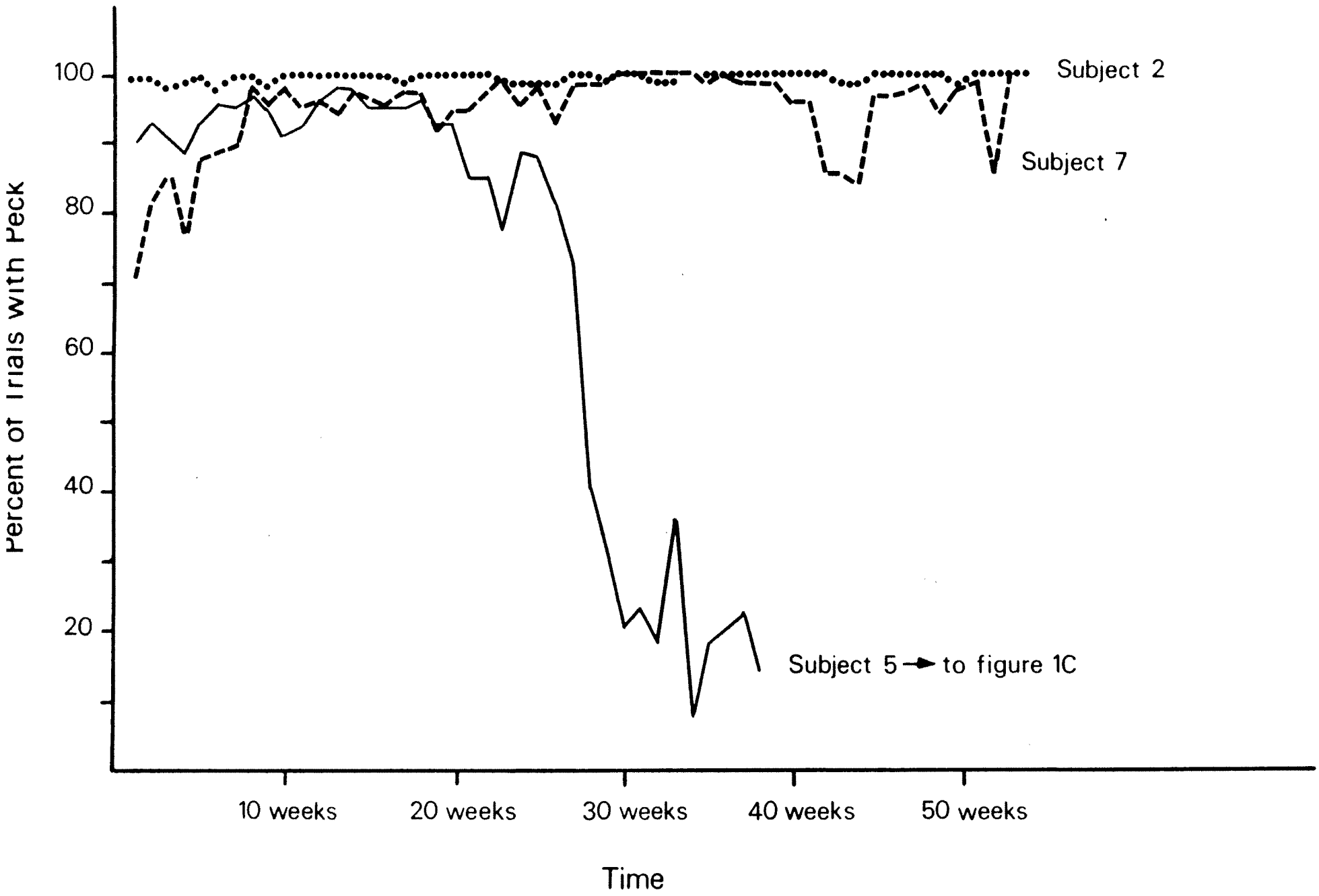
100
80
60
40
20
Percent of Trials with Peck
Subject 2
Subject 7
Subject 5 → to figure 1C
10 weeks
20 weeks
30 weeks
40 weeks
50 weeks
Time

Figure 1B

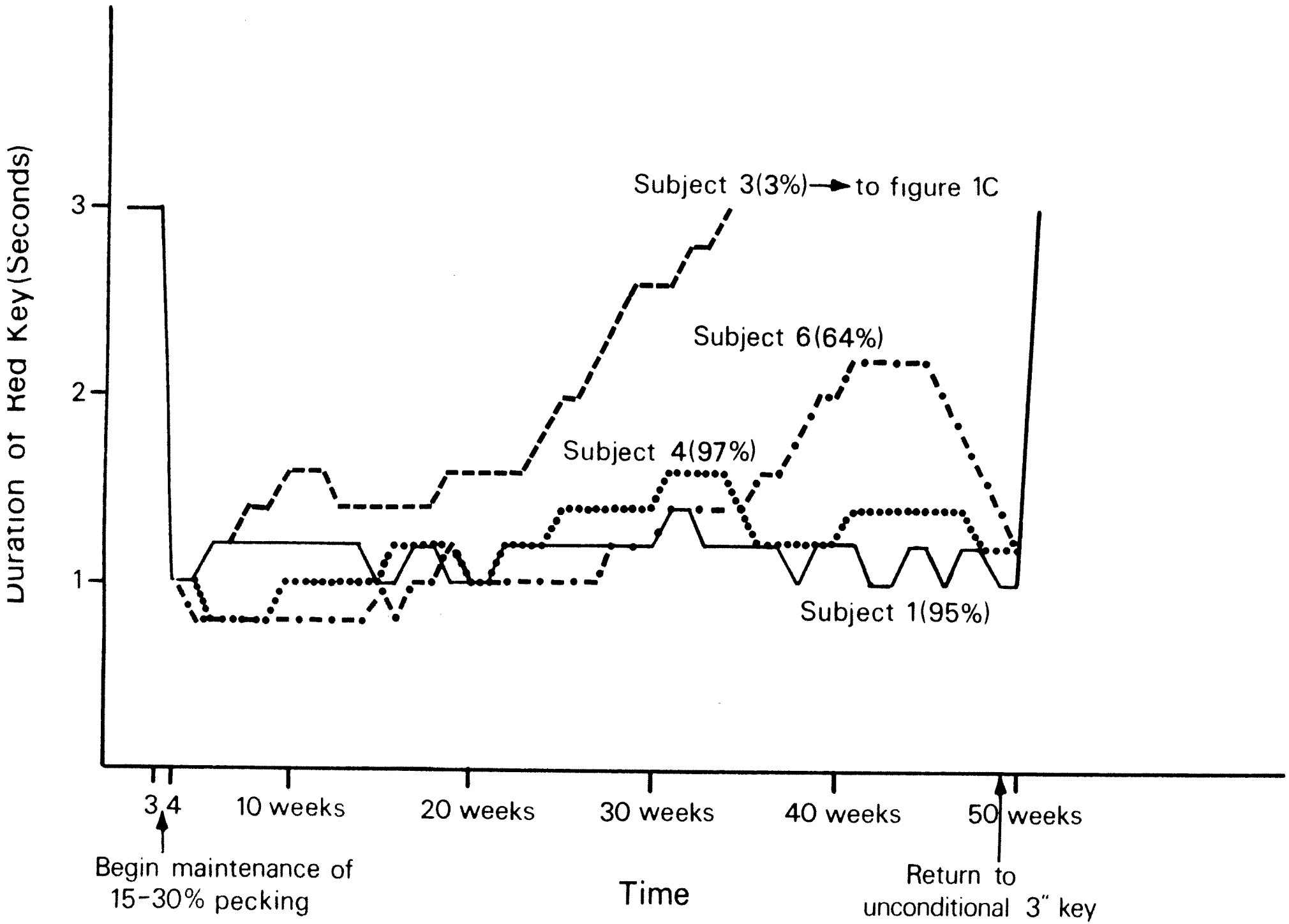
Duration of Red Key(Seconds)
3
2
1
Subject 3(3%)→to figure 1C
Subject 6(64%)
Subject 4(97%)
Subject 1(95%)
3,4
10 weeks
20 weeks
30 weeks
40 weeks
50 weeks
Begin maintenance of 15-30% pecking
Time
Return to unconditional 3" key

Figure 1C

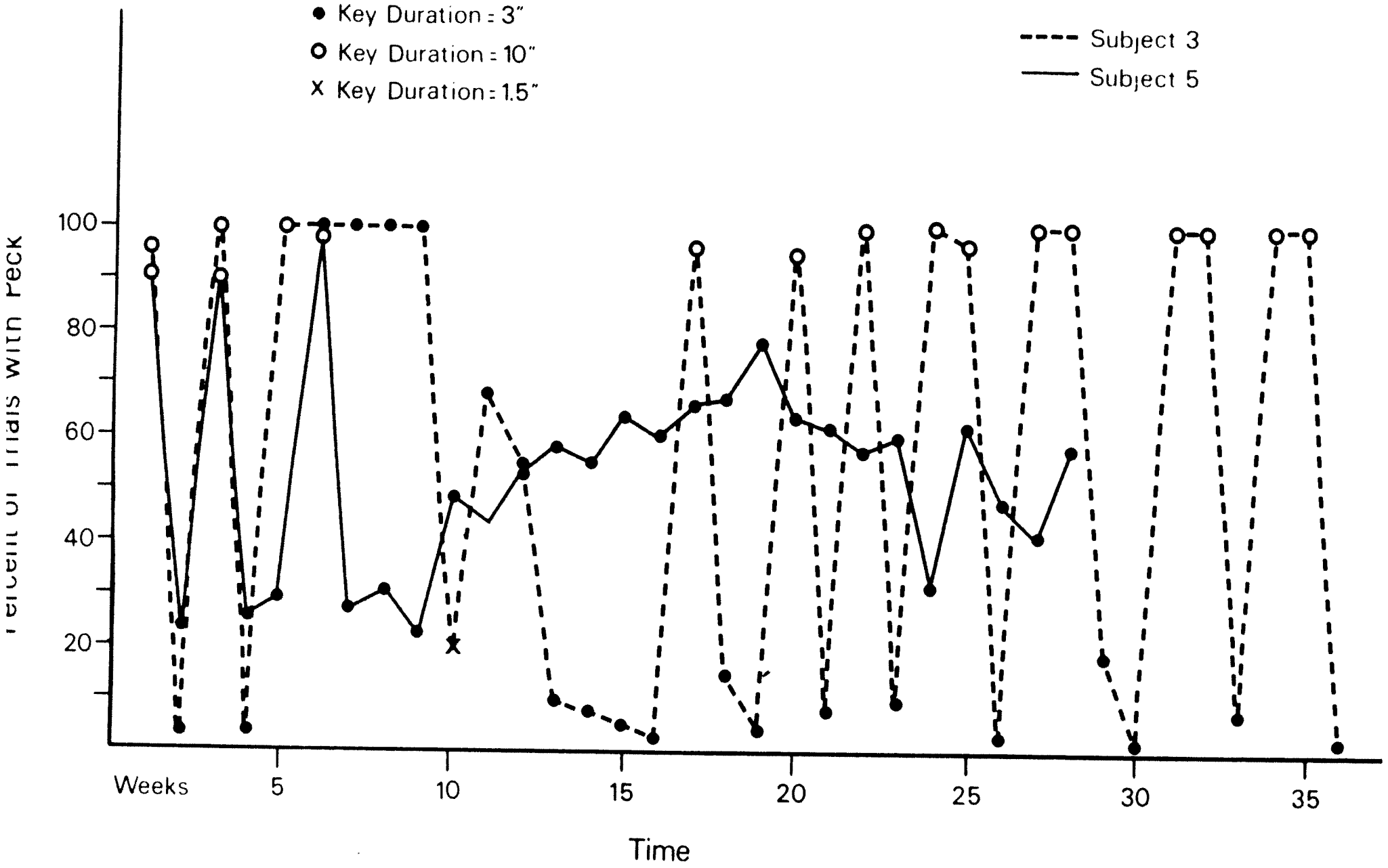
● Key Duration = 3"
○ Key Duration = 10"
X Key Duration = 1.5"
Subject 3
Subject 5
Percent of Trials with Peck
100
80
60
40
20
Weeks
5
10
15
20
25
30
35
Time

FIGURE 2 Birds 1-4

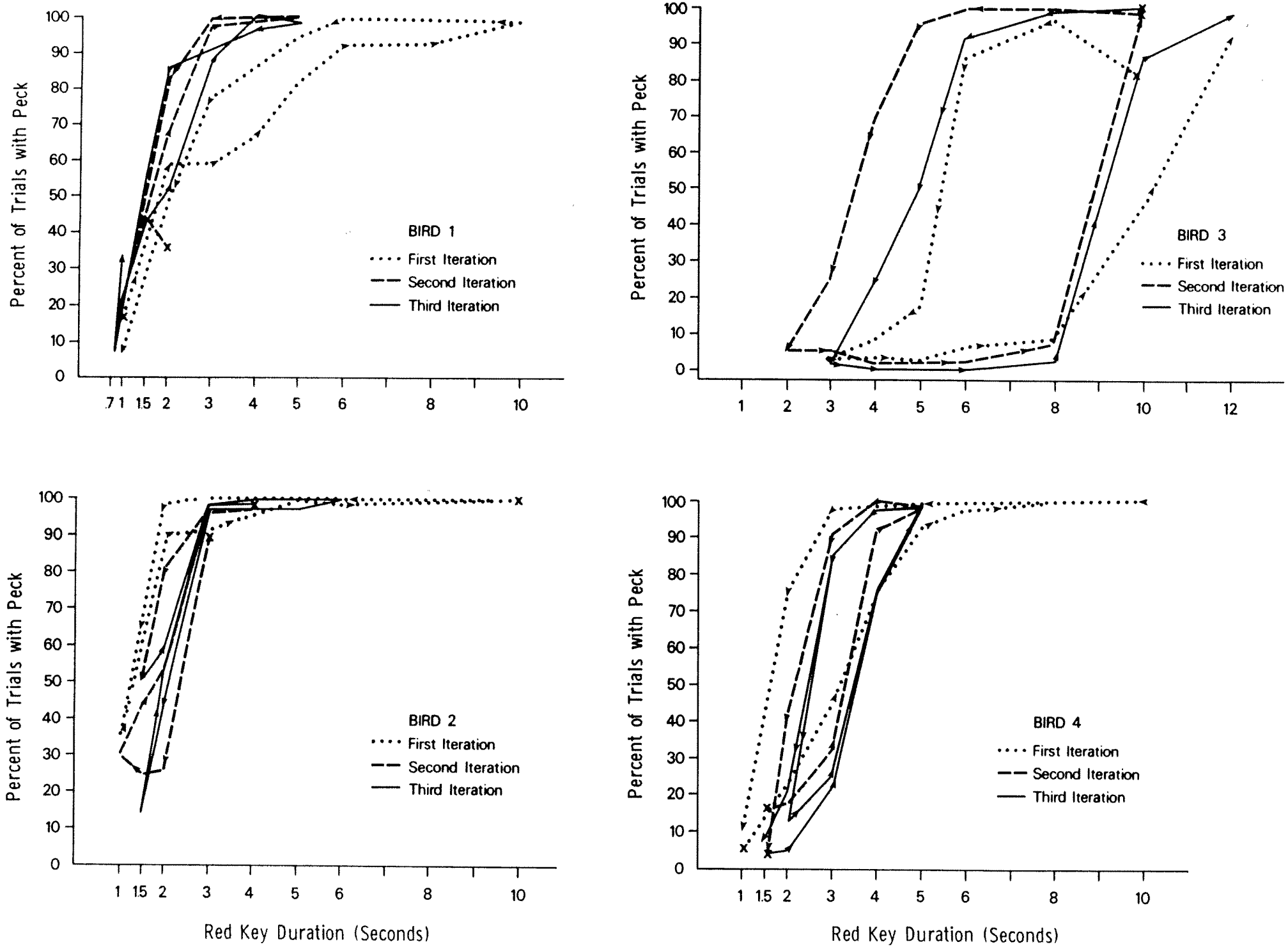
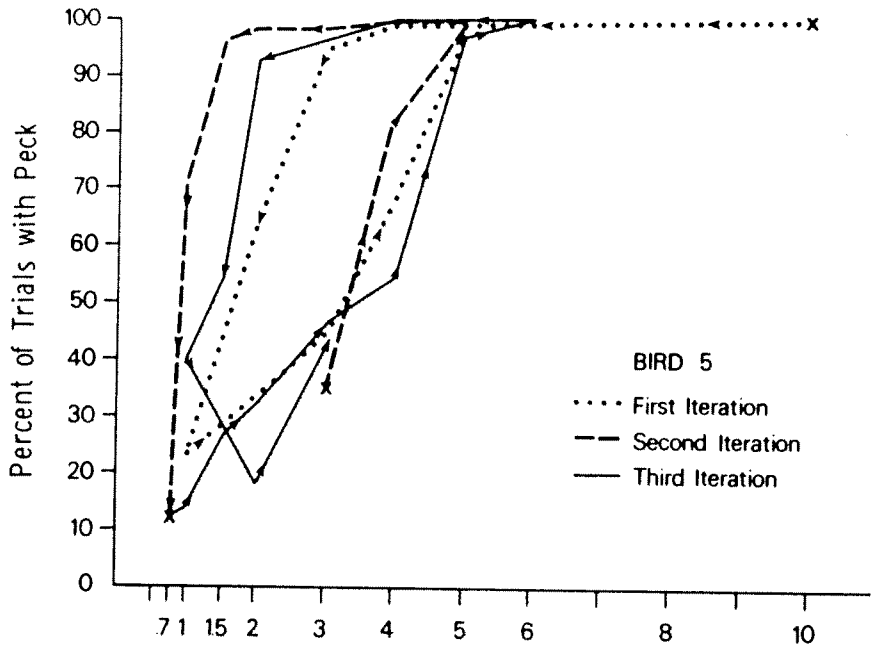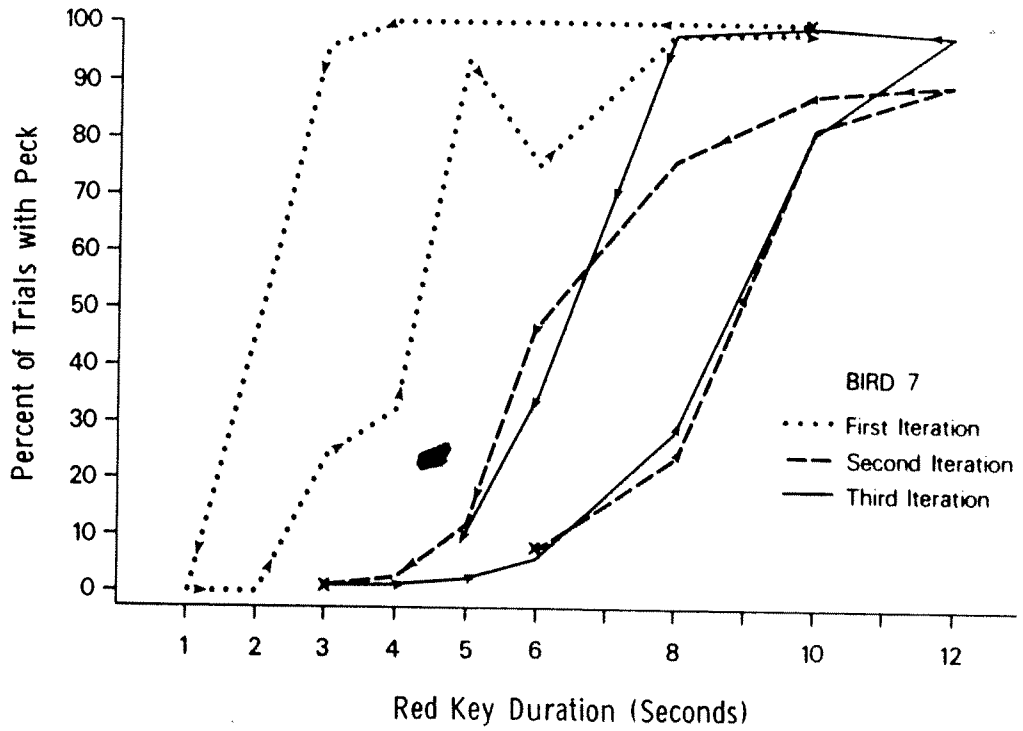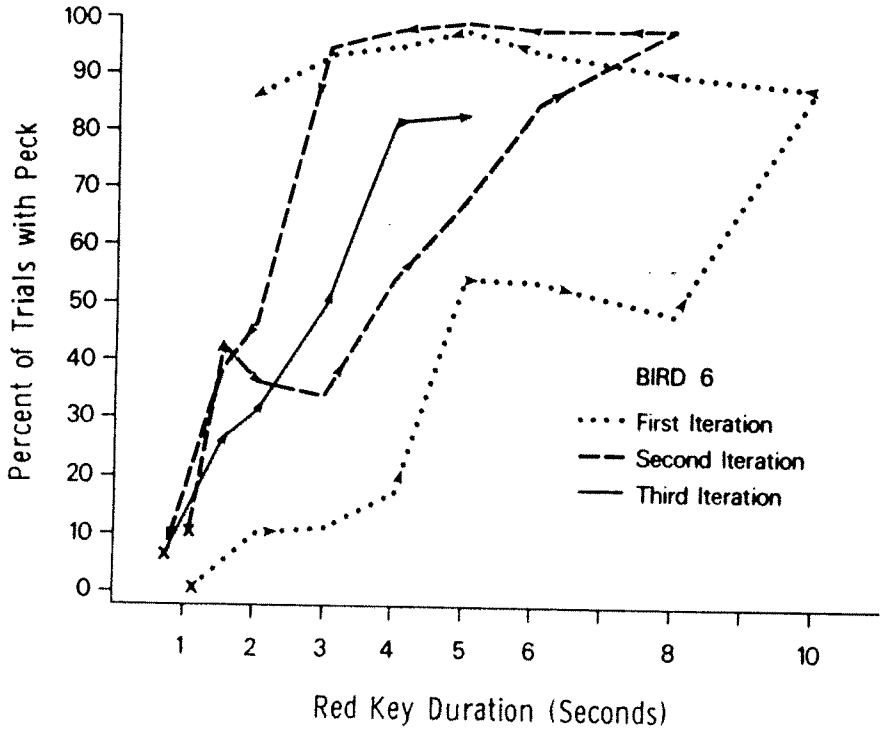
FIGURE 2 BIRDS 5-7