



Pigeons can learn a difficult discrimination if reinforcement is delayed following choice

Dalton House¹ · Daniel Peng¹ · Thomas R. Zentall¹

Received: 25 November 2019 / Revised: 15 January 2020 / Accepted: 20 January 2020 / Published online: 21 February 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Delaying reinforcement typically has been thought to retard the rate of acquisition of an association, but there is evidence that it may facilitate acquisition of some difficult simultaneous discriminations. After describing several cases in which delaying reinforcement can facilitate acquisition, we suggest that under conditions in which the magnitude of reinforcement is difficult to discriminate, the introduction of a delay between choice and reinforcement can facilitate the discrimination. In the present experiment, we tested the hypothesis that the discrimination between one pellet of food for choice of one alternative and two pellets of food for choice of another may be a difficult discrimination when choice consists of a single peck. If a 10-s delay occurs between choice and reinforcement, however, the discrimination is significantly easier. It is suggested that when discrimination between the outcomes of a choice is difficult and impulsive choice leads to immediate reinforcement, acquisition may be retarded. Under these conditions, the introduction of a brief delay may facilitate acquisition.

Keywords Discrimination · Magnitude of reinforcement · Delay of reinforcement · Pigeons

Introduction

The subjective value of reinforcement depends in part on the delay to reinforcement (Hull 1943). Typically, the longer the delay, the lesser the value. The effect of delay of reinforcement can be seen in the delay discounting effect in which a smaller reinforcer sooner may be preferred over a larger reinforcer later (Ainslie 1974). The delay discounting function describes the degree to which an organism will wait for the larger delayed reinforcement as the delay increases. The slope of the delay discounting function (reflecting how little time must pass before the organism switches from the larger later reinforcer to the smaller sooner reinforcer) reflects the impulsivity of the organism or its lack of self-control.

In a classic paper, Rachlin and Green (1972; see also Kurth-Nelson and Redish 2012) found that pigeons would choose the larger later more often if they would make a prior “commitment”. To accomplish this, at a time before the choice of smaller sooner versus larger later, the pigeons were allowed to choose between having a delayed choice

and being forced to take the larger later without a choice. Interestingly, the pigeons preferred choosing not to have a later choice. One could describe this initial choice as not being “tempted” later to make the smaller sooner choice. Although this choice not to choose may seem paradoxical, it actually is predicted from the effect of the delay, in delay discounting research. If other variables are held constant, the preference between two intervals will depend on the ratio of the two intervals. For example, if the delay to the smaller sooner is 1 s and the delay to the larger later is 10 s, then the ratio of the two would be 1:10, a ratio that strongly favors the sooner reinforcement. On the other hand, if the initial choice is made 10 s earlier, then the ratio of the two would be 11:20 or about 1:2, a ratio that may be overcome by the delayed larger reinforcer. The idea that the temporal difference between two reinforcers can be reduced subjectively by increasing the delay to both reinforcers is well understood but what if the difference in the magnitude of reinforcement is not so great (2:1 rather than 4:1) such that the discrimination between them is difficult and the temporal difference between two reinforcers is small or nonexistent? Under these conditions, the effect of a joint delay of reinforcement of the choice response for both alternatives is not so clear. It is also possible that by inserting a delay between smaller reinforcer sooner it

✉ Thomas R. Zentall
zentall@uky.edu

¹ Department of Psychology, University of Kentucky,
Lexington, KY 40506, USA

reduces impulsive choice resulting from the relative immediacy of reinforcement.

Consider the ephemeral reward task, a curious task that was first studied by Bshary and Grutter (2002). With this task, an animal is given a choice between two seemingly equal reinforcers, A and B. If the animal chooses A, it obtains the reinforcer and the trial is over. If it chooses B, however, it obtains the reinforcer but A remains and it can have that reinforcer as well. Thus, simply put, choice of A provides one reinforcer, whereas choice of B provides two. Strangely, with this task, although cleaner fish (wrasse; Salwiczek et al. 2012) and parrots (Pepperberg and Hartsfield 2014) quickly learn to choose B, the optimal choice, most primates (Salwiczek et al. 2012), rats (Zentall et al. 2017c), and pigeons (Zentall et al. 2016) have great difficulty learning to choose optimally. Although in the ephemeral reward task, several species appear to have great difficulty in associating the second reinforcer with their choice, Zentall et al. (2017b, c) found that both rats and pigeons can learn to choose optimally, if a delay is inserted between the choice and the first reinforcer. These results suggest that in the delay discounting procedure, the temporal ratio between the two alternatives and the difference in the magnitude of reinforcement may not be the only mechanisms responsible for the choice between the smaller sooner and the larger later.

In the ephemeral reward task, to some animals, the task may appear to involve a choice between two equal reinforcers, but sometimes there is a second reinforcer and for some reason the second reinforcer is not associated with the initial choice. Because the initial choice appears to involve a choice between two immediate reinforcers, one might consider the indifference between the two options as a form of impulsivity.

The results of recent research on object permanence in pigeons offers support for this hypothesis. Zentall and Raley (2019) studied object permanence in pigeons using a two-alternative choice apparatus. After a small amount of grain was placed into one of two cups, the pigeon was offered a choice between the cups. Surprisingly, the pigeons did not consistently choose the cup with the food. Furthermore, the pigeon did not even appear to be able to *learn* to choose the cup with the grain. Certainly, one would think that the pigeon should have been able to learn to associate the hand baiting the cup with the cup and the grain (but of course such a finding would not be evidence for object permanence). Observation of the pigeons, however, suggested that their high degree of motivation at the time of baiting might have caused the pigeons to choose the first cup that they saw, very likely an impulsive choice. In support of this hypothesis, in a follow-up experiment, we found that if we waited a short time (5 s) after baiting the cup, the pigeons' choice was much more accurate.

Another example of a delay-facilitated outcome discrimination occurred in a suboptimal choice task that has been compared to human unskilled gambling. When pigeons are given a choice between 20% signaled reinforcement and 50% unsignaled reinforcement, they show a strong preference for the suboptimal 20% reinforcement (Stagner and Zentall 2010). If a delay is inserted between choice and the conditioned reinforcer, however, they make significantly fewer suboptimal choices (Zentall et al. 2017a; see also McDevitt et al. 1997).

Although delay of reinforcement has generally been thought to retard acquisition (Renner 1964), when it comes to the acquisition of simultaneous discriminations that have outcomes that are difficult to discriminate, under certain conditions, inserting a delay between choice and reinforcement may facilitate learning. In the present research, the delay to reinforcement will be defined as the delay between the choice response and reinforcement.

In delay discounting research, one often starts with a discrimination between two stimuli signaling two different magnitudes of reinforcement without a delay, to demonstrate that the organism can discriminate between them and choose the larger one. There is evidence that pigeons can readily discriminate between stimuli signaling one versus four pellets at equal probabilities and delays (e.g., Smith et al. 2017). Under similar conditions, in unpublished research, we have found, however, that pigeons have great difficulty discriminating between stimuli signaling one versus two pellets. Although it is possible that the difference in reinforcement magnitude between one and two pellets is insufficiently large for pigeons to discriminate, it is also possible that impulsive choice interferes with acquisition of the discrimination. If this is so, it may be that inserting a delay between choice and the reinforcer might reduce impulsivity and facilitate the discrimination.

Thus, the purpose of the present experiment was to test the hypothesis that acquisition of a difficult magnitude-of-reinforcement discrimination might be facilitated by delaying the time between choice and reinforcement. In this experiment, we compared acquisition of a discrimination between one and two pellets of reinforcement. For pigeons in the Control group, the choice response was defined as a single peck to a red or green light. For pigeons in the Experimental group, the choice response was a choice peck to a red or a green light that initiated a fixed interval 10-s schedule (the first peck after 10 s provided reinforcement).

Method

Subjects

The subjects were 12 unsexed white Carneau pigeons, 5–8 years old, obtained from the Palmetto Pigeon Plant

(Sumter, South Carolina, USA). All of the pigeons had prior experience with simultaneous color discriminations in a midsession reversal experiment in which pecking to one color was reinforced for the first 40 trials of each session and pecking to the other color was reinforced for the remaining 40 trials of each session. All of the pigeons had free access to water and grit in a climate-controlled colony room that was maintained on a 12:12-h light/dark cycle. All testing occurred during the light cycle (7 a.m.–7 p.m.). During the experiment, all pigeons were maintained at 85% of their free-feeding body weight and were cared for in accordance with the University of Kentucky's Animal Care Guidelines.

Apparatus

All sessions were conducted in a Med Associates (St Albans, VT) ENV-008 modular operant test chamber (53.3 cm from the response panel to the back wall \times 34.9 cm across the response panel \times 12.0 cm high). The response panel in the chamber had a horizontal row of three response keys. Behind each key was a 12-stimulus inline projector (Industrial Electronics Engineering, Van Nuys, CA) that projected red, and green hues (Kodak Wratten Filter Nos. 2 and 60, respectively). Reinforcement was delivered by a pellet dispenser (Med Associates ENV-203-45) mounted behind the response panel that dispensed 45-mg pellets to a centrally located pellet tray. Only the left and right response keys were used. The apparatus was located in a sound attenuating chamber. A microcomputer in the adjacent room controlled the sessions by means of a Med Associates interface.

Procedure

Pretraining

All of the pigeons were pretrained to peck the red and green stimuli on the two side keys. For half of the pigeons ($n=6$), a single peck to the red side key resulted in one pellet of reinforcement and a peck to the green side key resulted in two pellets of reinforcement. For the remaining pigeons ($n=6$), a single peck to the green side key resulted in one pellet of reinforcement and a peck to the red side key resulted in two pellets of reinforcement. Each reinforcement advanced to the next trial after a 10-s intertrial interval, during which the house light was lit. The pigeons received ten presentations of each stimulus on the left and right side keys.

For the second pretraining session, a random half of the pigeons in each group were assigned to the Experimental group for which ten pecks were required for each of the red and green stimuli for reinforcement. The remaining pigeons were assigned to the Control group for which a single peck was required for each of the stimuli for reinforcement.

Training

The first six training sessions consisted of 40 choice trials in which a red stimulus was presented on one side key and a green stimulus was presented on the other. The location of the stimuli (left or right) was counterbalanced over subjects. For the Experimental group, the first peck turned off the key not chosen and initiated a fixed interval 10-s schedule (the first response after 10 s resulted in reinforcement). We switched the pigeons in the Experimental group to a fixed interval schedule to better standardize trial duration. For the Control group, a single peck was required for reinforcement. After six sessions of training, the pigeons showed strong position preferences. In an attempt to break those preferences, we replaced 20 choice trials with forced trials in which a single stimulus was presented on one of the side keys (counterbalanced for color and location) with the same contingencies in effect as on choice trials. The forced trials ensured that the pigeons pecked each side key and color on at least ten trials in each session. The remaining 20 trials in each session were choice trials. Training with a mixture of forced and choice trials consisted of 22 40-trial sessions.

Results

All pigeons began training at close to 50% choice of the optimal (two-pellet) alternative. The pigeons in the Experimental group, the group for which there was a delay between choice and reinforcement, showed a regular increase in their choice of the optimal alternative, reaching 80% optimal choice within 20 sessions of training (see Fig. 1). The pigeons in the Control group, the group for which only a single peck to the chosen alternative was required, showed a relatively small increment in their choice of the optimal alternative during training.

Analysis of the training data for the Experiment group, over the last five sessions of training, indicated that choice of the optimal alternative (83.4%) was significantly above chance, $t(5)=5.38$, $p=0.003$. Analysis of the training data for the Control group, over the last five sessions of training, indicated that the choice of the optimal alternative (57.0%) was not significantly above chance, $t(5)=0.84$, $p=0.44$. Analysis of the difference in choice of the optimal alternative between the Experiment group and the Control group, over the last five sessions of training, indicated that the Experimental group chose optimally significantly more than the Control group, $t(10)=2.55$, $p=0.027$.

Discussion

The purpose of the present experiment was to test the hypothesis that acquisition of a simple simultaneous discrimination in which the differential outcomes (one vs. two

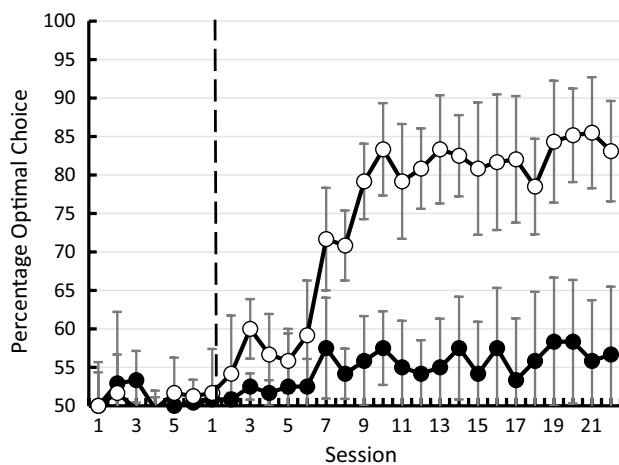


Fig. 1 Percentage choice of the optimal, two-pellet alternative over the suboptimal one-pellet alternative as a function of training session. For the Experimental group, the first peck turned off the key not chosen and initiated a fixed interval 10-s schedule. For the Control group, only one peck was required for each choice. For the first 6 sessions, all 40 trials in each session were choice trials. For the next 22 sessions, there were 20 forced trials and 20 choice trials in each session. Error bar \pm one SEM

pellets) that were likely to make the discrimination difficult, would be facilitated by delaying the outcome following choice. Pigeons in the Experiment group had to wait for 10 s following choice of the discriminative stimulus before the outcome was obtained, whereas pigeons in the Control group received reinforcement immediately following choice of the discriminative stimulus.

Delay of reinforcement generally has been thought to retard learning but there is now growing evidence that when outcomes are somewhat difficult to discriminate, inserting a delay between choice and reinforcement can facilitate acquisition. This phenomenon was first demonstrated in research on the ephemeral reward task, a task most similar the present task (Zentall et al. 2016). In the ephemeral reward task, pigeons were given a choice between two alternatives. Although the alternatives appeared to provide the same amount of food, choice of one ended the trial, whereas choice of the other allowed the pigeon to obtain the other reinforcer as well. When there was no delay between choice and the first reinforcement, pigeons were not able to learn to choose optimally. When a 20-s delay occurred between choice and the first reinforcement, pigeons learned to choose optimally.

Delaying the outcome of a choice also can result in less suboptimal choice in other contexts. For example, in the gambling analog task, pigeons show a preference for one alternative that provides a reliable cue for reinforcement, but only 20% of the time, over the other alternative that always provides a cue for 50% reinforcement (Stagner and Zentall 2010). When a delay is inserted between choice and the cues

for reinforcement that follow, there is a significant reduction in suboptimal choice (Zentall et al. 2017a).

A third example of a delay that facilitates acquisition is in an object permanence procedure (Zentall and Raley 2019). When grain is dropped into one of two cups in front of a pigeon, the pigeon appears to have difficulty locating the cup with the grain. If, however, the experimenter waits for 5 s before allowing the pigeon access to the cups, the pigeon learns which cup has the grain. Furthermore, it then can choose appropriately following the delay, even if the two cups are rotated 90° or even 180° after baiting.

The results of the present experiment are different from those involving the commitment procedure in delay discounting (Rachlin and Green 1972; Siegel and Rachlin 1995). In delay discounting, organisms readily discriminate between the larger and smaller reinforcer when the larger reinforcer is not delayed. In the present procedure, however, the difference between the larger two pellets and the smaller one pellet is small and the pigeons do not appear to discriminate between them unless a delay is inserted between the choice response and reinforcement. Thus, the present procedure does not involve differential delay between the two reinforcers as in the commitment research.

In the present procedure, the rate of reinforcement between the Experimental and Control groups was not the same. For the Control group, the time between reinforcers was approximately 11 s (about 1 s for choice and 10 s for the intertrial interval). For the Experimental group, however, the time between reinforcers was approximately 21 s (about 1 s for choice, 10 s for the fixed interval schedule, and 10 s for the intertrial interval). Is it possible that the reduced rate of reinforcement associated with the Experimental group (about 2.86 rf/min for the Experimental group vs. 5.45 rf/min for the Control group) could account for facilitated acquisition by the pigeons in the Experimental group? That is, would we have obtained the same result had we increased the intertrial interval for the Control group to 20 s?

There is evidence, for example, that in Pavlovian procedures, conditioning is better, the longer the intertrial interval or time between unconditioned stimuli (see e.g., Holland 2000). In choice procedures, however, typically the effect of the time between trials has been found to have little effect on choice accuracy. For example, Rayburn-Reeves et al. (2013) found that in a simple simultaneous discrimination there was little effect on discrimination accuracy when the intertrial interval was varied between 1.5 and 10 s between groups. Similarly, Grossi (1981) found that in a successive discrimination, the discrimination index was unaffected by manipulation of the intertrial interval between 1 and 25 s. Similar manipulation of the intertrial interval in a conditional discrimination between 5 and 25 s was found to have little effect on conditional discrimination accuracy (Santi 1984). Thus, it is unlikely

that for the Experimental group, the reduced rate of reinforcement resulting from the increased time between reinforcers was responsible for the improved accuracy for that group.

In addition to the time between reinforcements, the pigeons in the Experimental group made more key pecks than the Control group. However, previous research that demonstrated an effect of delay between choice and reinforcement but controlled for both responses made and time between reinforcers showed a similar effect to that found in the present experiment (Zentall et al. 2017a).

The results of the present experiment confirm and extend the results of the earlier research. Paradoxically, although typically delay of reinforcement has been thought to retard learning, in simultaneous discriminations with difficult to discriminate outcomes, it actually appears to facilitate learning. The mechanism responsible for this facilitated learning by inserting a delay between choice and reinforcement is not obvious, but the results of inserting a delay following baiting in the object permanence task may provide a working hypothesis. When one of two cups is baited with grain, it is presumed that the pigeon can see and hear the grain being placed in the cup so why would the pigeon ever choose the unbaited cup. It is possible that the sight and sound of the grain impulsively causes the pigeon to choose the closest cup. Furthermore, the reinforcement for impulsive choice on 50% of the trials may be sufficient to maintain the impulsivity. Although the 5-s delay may allow time for the pigeon to forget which cup was baited, it also may decrease the likelihood of it making an impulsive choice. In the present experiment, as well, the immediacy of reinforcement for choice of either stimulus may encourage impulsive choice. Thus, the 5-s delay between choice and reinforcement may be sufficient to reduce impulsive choice and facilitate optimal choice.

The reduction in the likelihood of impulsive choice may also account for the effect of delay in the ephemeral reward experiment as well as in the present experiment. It is very likely that the immediacy of reinforcement for impulsive choice maintains that behavior and in the case of the ephemeral reward task, blocks the association of the choice with the second reinforcer that follows. If impulsivity is the source of suboptimal choice, it suggests that the immediacy of reinforcement results in a performance problem rather than a learning problem.

In the case of the gambling analog experiments, the added delay between choice and signal for reinforcement may also reduce the likelihood of impulsive choice. In the gambling analog experiments, the single peck sometimes results in a conditioned reinforcer that always predicts reinforcement. Adding a delay between choice and the appearance of the conditioned reinforcer that always predicts reinforcement is thought to have an effect similar to the effect in the present

experiment in which the delay was between choice and the reinforcer.

The present results may have implications for other contexts in which the immediacy of reinforcement (or reliable signals for reinforcement) following a choice response may result in impulsive responding that leads to suboptimal choice. Delaying the consequences of such impulsive choice may lead to less suboptimal choice and may result in better overall accuracy.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

References

- Ainslie GW (1974) Impulse control in pigeons. *J Exp Anal Behav* 21:485–489
- Bshary R, Grutter AS (2002) Experimental evidence that partner choice is a driving force in the payoff distribution among cooperators or mutualists: the cleaner fish case. *Ecol Lett* 5:130–136
- Grossi V (1981) Pigeon short-term memory: the effects of intertrial interval duration and illumination in a successive delayed matching-to-sample procedure. Unpublished master's thesis, Wilfrid Laurier University, Waterloo, Ontario, Canada
- Holland PC (2000) Trial and intertrial durations in appetitive conditioning in rats. *Anim Learn Behav* 28:121–135
- Hull CL (1943) Principles of behavior. Appleton-Century-Crofts, New York
- Kurth-Nelson Z, Redish AD (2012) Don't let me do that!—models of precommitment. *Front Neurosci* 6:138. <https://doi.org/10.3389/fnins.2012.00138>
- McDevitt MA, Spetch ML, Dunn R (1997) Contiguity and conditioned reinforcement in probabilistic choice. *J Exp Anal Behav* 68(3):317–327. <https://doi.org/10.1901/jeab.1997.68-317>
- Pepperberg IM, Hartsfield LA (2014) Can Grey parrots (*Psittacus erithacus*) succeed on a “complex” foraging task failed by non-human primates (*Pan troglodytes*, *Pongo abelii*, *Sapajus apella*) but solved by wrasse fish (*Labroides dimidiatus*)? *J Comp Psychol* 128:298–306
- Rachlin H, Green L (1972) Commitment, choice and self-control. *J Exp Anal Behav* 17:15–22
- Rayburn-Reeves RM, Laude JR, Zentall TR (2013) Pigeons show near-optimal win-stay/lose-shift performance on a simultaneous-discrimination, midsession reversal task with short intertrial intervals. *Behav Proc* 92:65–70. <https://doi.org/10.1016/j.beproc.2012.10.011>
- Renner KE (1964) Delay of reinforcement: a historical view. *Psychol Bull* 61:341–361
- Salwiczek LH, Prétôt L, Demarta L, Proctor D, Essler J, Pinto AI, Wismer S, Stoinski T, Brosnan SF, Bshary R (2012) Adult cleaner wrasse outperform capuchin monkeys, chimpanzees, and orangutans in a complex foraging task derived from cleaner–client reef fish cooperation. *PLoS ONE* 7:e49068. <https://doi.org/10.1371/journal.pone.0049068>

- Santi A (1984) The trial spacing effect in delayed matching-to-sample by pigeons is dependent upon the illumination condition during the intertrial interval. *Can J Psychol* 38:154–165
- Siegel E, Rachlin H (1995) Soft commitment: self-control achieved by response persistence. *J Exp Anal Behav* 64:117–128
- Smith AP, Beckmann JS, Zentall TR (2017) Gambling-like behavior in pigeons: ‘jackpot’ signals promote maladaptive risky choice. *Nat Sci Rep* 7:6625. <https://doi.org/10.1038/s41598-017-06641-x>
- Stagner JP, Zentall TR (2010) Suboptimal choice behavior by pigeons. *Psychon Bull Rev* 17:412–416
- Zentall TR, Raley OL (2019) Object permanence in the pigeon: insertion of a delay prior to choice facilitates visible- and invisible-displacement accuracy. *J Comp Psychol* 133:132–139
- Zentall TR, Case JP, Luong J (2016) Pigeon’s paradoxical preference for the suboptimal alternative in a complex foraging task. *J Comp Psychol* 130:138–144
- Zentall TR, Andrews DM, Case JP (2017a) Prior commitment: its effect on suboptimal choice in a gambling-like task. *Behav Proc* 145:1–9
- Zentall TR, Case JP, Berry JR (2017b) Early commitment facilitates optimal choice by pigeons. *Psychon Bull Rev* 24(3):957–963
- Zentall TR, Case JP, Berry JR (2017c) Rats’ acquisition of the ephemeral reward task. *Anim Cogn* 20:419–425

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.