

# The Paradoxical Effect of Low Reward Probabilities in Suboptimal Choice

Inês Fortes, Carlos Pinto, and Armando Machado  
University of Minho

Marco Vasconcelos  
University of Aveiro and CESAM, Centre for Environmental  
and Marine Studies

When offered a choice between 2 alternatives, animals sometimes prefer the option yielding less food. For instance, pigeons and starlings prefer an option that on 20% of the trials presents a stimulus always followed by food, and on the remaining 80% of the trials presents a stimulus never followed by food (the Informative Option), over an option that provides food on 50% of the trials regardless of the stimulus presented (the Noninformative Option). To explain this suboptimal behavior, it has been hypothesized that animals ignore (or do not engage with) the stimulus that is never followed by food in the Informative Option. To assess when pigeons attend to the stimulus usually not followed by food, we increased the probability of reinforcement,  $p$ , in the presence of that stimulus. Across 2 experiments, we found that the value of the Informative Option decreased with  $p$ . To account for the results, we added to the Reinforcement Rate Model (and also to the Hyperbolic Discounting Model) an engagement function,  $f(p)$ , that specified the likelihood the animal attends to a stimulus followed by reward with probability  $p$ , and then derived the model predictions for 2 forms of  $f(p)$ , a linear function, and an all-or-none threshold function. Both models predicted the observed findings with a linear engagement function: The higher the probability of reinforcement after a stimulus, the higher the probability of engaging the stimulus, and, surprisingly, the less the value of the option comprising the stimulus.

**Keywords:** suboptimal choice, engagement, reinforcement, Reinforcement Rate Model, pigeons

Animals attend to important stimuli and ignore unimportant ones. When foraging, a squirrel attends to the dangerous eagle above but ignores the harmless rabbit nearby. In the laboratory, there is also ample evidence that stimuli that signal important events (e.g., food to a hungry animal, electric shock) are attended more than less relevant stimuli. Not only is attending selective, but it is also dynamic, for animals may initially attend to but later ignore an event, and vice versa. A naïve bird may initially freeze to the sound of the experimenter coming into the room, but later ignores the same sound.

Because attending to important events is adaptive, seeming violations of the rule deserve scrutiny. By studying tasks in which an animal seems to ignore important events, we may discover the

constraints of its behavioral processes. This article focus on one such task with pigeons, and it attempts to explain the pigeons' seemingly maladaptive behavior. Stagner and Zentall (2010; also Stagner, Laude, & Zentall, 2012; Zentall & Stagner, 2011b) exposed pigeons to a concurrent-chain schedule with two options in the initial link (see Figure 1): the Informative Option (left white key) and the Noninformative Option (right white key). If the pigeon chooses the Informative Option, the left white key changes to either Red (with probability 0.2) or Green (with probability 0.8). After a delay of 10 s, if the key is Red the pigeon receives food, but if the key is Green the pigeon does not receive food. We represent these contingencies by  $S_{\text{Red},1.0}$  (i.e., Red signals that food will always occur) and  $S_{\text{Green},0.0}$  (i.e., Green signals the absence of food). Hence, the two colors in the left terminal link signal unambiguously the trial outcome.

If the pigeon chooses the Noninformative Option, the right white key changes to either Yellow (with probability 0.2) or Blue (with probability 0.8). After a delay of 10 s the pigeon receives food with probability 0.5 regardless of the key color. We represent these contingencies by  $S_{\text{Yellow},0.5}$  and  $S_{\text{Blue},0.5}$ . Hence, the two colors in the right terminal link are unreliable predictors of the trial outcome.

From the foregoing probabilities, we calculate the probability of food in the Informative Option as 0.2 ( $0.2 \times 1 + 0.8 \times 0$ ) and in the Noninformative Option as 0.5 ( $0.2 \times 0.5 + 0.8 \times 0.5$ ). To maximize reinforcement rate (overall or local), the pigeons should prefer the Noninformative Option. Surprisingly, however, pigeons and starlings strongly prefer the Informative Option (e.g., Stagner et al., 2012; Stagner & Zentall, 2010; Vasconcelos, Monteiro, & Kacelnik, 2015; Zentall, 2016; Zentall & Stagner, 2011b). They choose as if they ignored the consequences associated with the

Inês Fortes, Carlos Pinto, and Armando Machado, School of Psychology, University of Minho; Marco Vasconcelos, Department of Education and Psychology, and Department of Biology, University of Aveiro and CESAM, Centre for Environmental and Marine Studies, Portugal.

This study was conducted at the Psychology Research Centre (UID/PSI/01662/2013) of the University of Minho, and supported by the Portuguese Foundation for Science and Technology (FCT) and the Portuguese Ministry of Science, Technology and Higher Education through national funds and cofinanced by FEDER through COMPETE2020 under the PT2020 Partnership Agreement (POCI-01-0145-FEDER-007653). This work was also supported by an FCT Grant (PTDC/MHC-PCN/3540/2012) to AM. IF and CP were supported by FCT Doctoral Grants (SFRH/BD/77061/2011 and SFRH/BD/78566/2011, respectively).

Correspondence concerning this article should be addressed to Inês Fortes, School of Psychology, University of Minho, Campus de Gualtar, 4710-057 Braga. E-mail: [ines.fortes@gmail.com](mailto:ines.fortes@gmail.com)

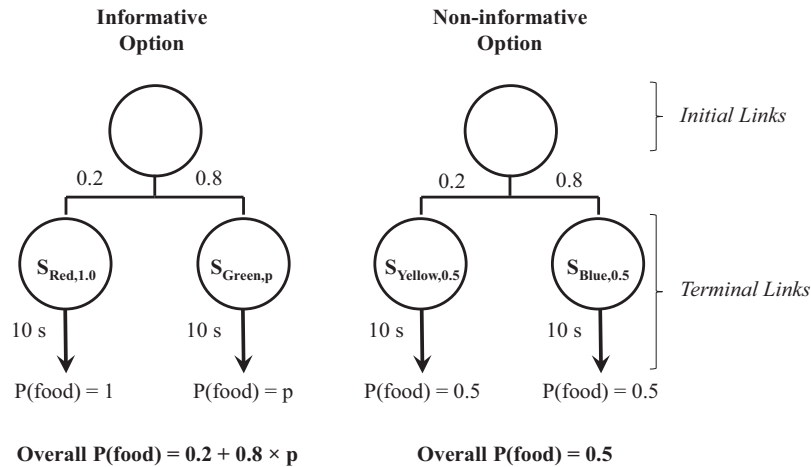


Figure 1. Design of Experiment 1. The probability of reinforcement following  $S_{Green,p}$  varied across groups from  $p = 0$  to  $p = .375$ .

Informative Option, the lower overall and local reinforcement rates.

To explain performance in this task, several models (e.g., Reinforcement Rate Model [RRM], Fortes, Machado, & Vasconcelos, 2017; Fortes, Vasconcelos, & Machado, 2016; Vasconcelos et al., 2015; and the Hyperbolic Discounting Model [HDM], Mazur, 1984, 1987, 1997), assume that this suboptimal behavior occurs because pigeons *ignore the negative discriminative stimulus*,  $S_{Green,0.0}$ , and behave as if it was never presented. That is, pigeons perceive the Informative Option as always leading to  $S_{Red,1.0}$ , the Red keylight color followed by food, and the Noninformative Option as always leading to stimuli that are intermittently followed by food. As a result, the probability of reinforcement in the Informative Option is perceived as higher than that in the Noninformative Option.

Several experiments have confirmed this hypothesis by showing that changing the duration or probability of  $S_{Green,0.0}$  (originally 10 s and 0.8, respectively) has virtually no effect on choice (Fortes et al., 2016; Mazur, 1995, 1996; Pisklak, McDevitt, Dunn, & Spetch, 2015; Spetch, Mondloch, Belke, & Dunn, 1994; Zentall, Laude, Stagner, & Smith, 2015). The clearest case occurred when we increased the duration of  $S_{Green,0.0}$  from 10 to 200 s; thus, decreasing drastically the reinforcement rate in the Informative Option (Fortes et al., 2016, Experiment 2). Even at the longest delays, the pigeons preferred the Informative Option on 95% of the trials, consistent with the hypothesis that pigeons do not take into account the  $S_{Green,0.0}$  durations.

If animals cease to pay attention to stimuli never followed by food ( $S_{Green,0.0}$ ), but pay attention to stimuli always or half of the times followed by food ( $S_{Red,1.0}$ ,  $S_{Blue,0.5}$ , and  $S_{Yellow,0.5}$ ), we may ask more generally how the probability of attending to a stimulus relates to the probability of food following that stimulus. To illustrate, suppose we increased the probability of food following  $S_{Green}$  from 0 to  $p$ . Would pigeons start attending to  $S_{Green}$  (gradually? abruptly?), and, if so, how would attending to  $S_{Green}$  affect their preference? The question is important because when an animal attends to a stimulus, it is affected not only by the rewards associated with the stimulus (the benefits of attending), but also by

the time spent waiting for the rewards (the costs of attending). Thus, attending to a stimulus means to engage with the stimulus and be affected by its positive and negative consequences. How these positive and negative consequences combine to determine the value of an alternative remains an unexplored topic.

Our aim is to examine how the value of the Informative Option changes as the probability of reinforcement,  $p$ , following  $S_{Green}$  increases. Given reasonable assumptions, varying  $p$  leads to counterintuitive predictions. For example, assume that animals (a) engage with  $S_{Green}$  provided  $p > 0$ , (b) choose the option with the highest *perceived* reinforcement rate, and (c) the latter equals to the ratio between the average probability of reinforcement associated with that option and the *perceived* duration of its terminal-link. For the Noninformative Option, the perceived reinforcement rate is independent of  $p$  and equals always 0.05 rewards/s (i.e.,  $0.5/(0.5 \times 10 + 0.5 \times 10)$ ). For the Informative Option, the perceived reinforcement rate depends on  $p$ : When  $p = 0$ , the animals do not attend to  $S_{Green}$  and, therefore, do not consider its duration. Hence, the perceived reinforcement rate of the Informative Option is 0.10 rewards/s (i.e.,  $0.2/(0.2 \times 10 + 0.8 \times 0)$ ). Because the rate is higher than the rate for the Noninformative Option, the animals should prefer the Informative Option. The prediction is consistent with the results obtained in the standard protocol.

However, when  $p$  increases to 0.1, the animal attends to  $S_{Green}$ , securing the benefits (i.e., the reinforcers) but also paying the costs (i.e., the delays), and the perceived reinforcement rate becomes 0.028 rewards/s (i.e.,  $(0.2 \times 1 + 0.8 \times 0.1)/(0.2 \times 10 + 0.8 \times 10)$ ). Because the rate is lower than the rate for the Noninformative Option, the animals ought to reverse preference. The prediction is counterintuitive because we would not expect preference to reverse when the probability of reinforcement for the preferred alternative increases.

Further increases in  $p$  increase the rate of the Informative Option. When  $p = .375$ , the perceived rate becomes 0.05 rewards/s, matching the rate of reinforcement in the Noninformative Option. Hence, for  $p > .375$ , preference should reverse again as the rate of the Informative Option again surpasses the rate of the

Noninformative Option. Whether or not animals pay attention to  $S_{\text{Green}}$  for any  $p > 0$  is at the moment uncertain.

We conducted two experiments to test the foregoing predictions and, more generally, to examine how preference changes with  $p$ . In Experiment 1, we varied  $p$  across phases and measured the relative value of the two options by the proportion of choices of the two initial links. In Experiment 2, we also varied  $p$  across phases but, within phases, we changed the delay in the terminal links of the Noninformative Option. We measured the relative value of the two options by the duration,  $D$ , that made the pigeons indifferent between the two initial links. We expect the experimental findings and theoretical analyses will bring new insights on how the probability of food following a stimulus affects the probability of attending to that stimulus.

## Experiment 1: Probability of Food on the Informative Option

### Method

**Subjects.** The subjects were eight pigeons (*Columba livia*) with previous experimental histories (mainly autoshaping and timing experiments). During the experiment, the animals were maintained at about 80% of their free-feeding weights. In the colony room, grit and water were continuously available. The pigeons were housed in a room with controlled temperature (between 20° and 22 °C) and light cycle (13h:11h light/dark, with lights on at 8 a.m.). All the animals were cared for according to the guidelines from the Portuguese Veterinary Authority (Direcção Geral de Alimentação e Veterinária).

**Apparatus.** Four Med Associates operant boxes for pigeons were used. The boxes were 28.5-cm high, 24-cm long, and 30-cm wide. Each box was enclosed in a sound-attenuating chamber, equipped with a fan that circulated air and masked extraneous noises. The response panel had three circular keys, 2.5 cm in diameter. The keys were 6 cm apart (center-to-center) and the lowest edge was 21 cm above the floor grid. Also in the response panel and 4 cm above the floor grid, a 6-cm wide  $\times$  5-cm high opening allowed the pigeon access to food. When the food hopper was raised, the opening was illuminated with a 1.1-W light in the panel opposite to the response panel a houselight (2.8 W) illuminated the whole box. This light was centrally located and 23 cm above the floor. A personal computer with a custom Visual Basic, 2010 program controlled the events and recorded data via Whisker software (Cardinal & Aitken, 2010).

### Procedure.

**Pretraining.** Two sessions of pretraining established pecking at all stimuli used in this task: red, green, yellow, and blue keylight colors on the left and right keys, and white on the left, center, and right keys. Each session had 44 trials, 4 trials with each stimulus. On each trial, one of the stimuli was presented, and one peck at the illuminated key was reinforced with 3 s of access to food, followed by a 10-s intertribal interval (ITI) during which the houselight was on. In the second session, 10 pecks to each stimulus were required for food reinforcement.

**Training.** The sessions comprised 120 trials, 40 forced-choice trials with each option (Informative and Noninformative) and 40 free-choice trials. The trials were randomly interspersed with the constraint that in each 30-trial block, 20 were forced-choice trials

(10 per option) and 10 were free-choice trials. For half of the birds, the Informative and Noninformative options were always presented on the left and right keys, respectively; for the other half, the side allocation was reversed.

A trial began with the white center key flashing (250 ms on, 250 ms off). On forced-choice trials, one peck at the center key turned it off and illuminated one of the side keys with a white hue (initial link, see Figure 1). One peck at the illuminated side key changed its color and initiated a 10-s delay. If the peck was at the Informative key, on 20% of the trials the key color changed to the  $S_{\text{Red},1.0}$  stimulus; after the 10-s delay the key turned off and food was delivered. On the remaining 80% of the trials, the key changed to the  $S_{\text{Green},p}$  stimulus and after the 10-s delay the key turned off and food was delivered with probability  $p$ . If the peck was at the Noninformative key, on 20% of the trials the key color changed to the  $S_{\text{Yellow},0.5}$  stimulus and on the other 80% of the trials the key color changed to  $S_{\text{Blue},0.5}$ . In both cases, the key stayed on for 10 s, after which food was delivered on 50% of the trials. The trials were separated by an illuminated 10-s ITI.

Free-choice trials were structurally similar to forced-choice trials, except that after a center-key peck, both side keys were illuminated; a peck to one of these keys turned the other key off, and changed the selected key to one of the terminal hues as in the forced-choice trials. The houselight was on only during the ITI.

The terminal-link hues associated with each option were counterbalanced across pigeons, but red and green were always associated with one option and blue and yellow were always associated with the other option. However, to simplify we describe the procedure and present the results as if all pigeons had the Green hue associated with  $p$ .

Thirty-two of the 40 Informative forced-choice trials (i.e., 80%) had the  $S_{\text{Green},p}$  stimulus as terminal link. The probabilities of reinforcement following  $S_{\text{Green},p}$  were manipulated between subjects and set at  $p = 0/32, 1/32, 6/32, \text{ or } 12/32$ . For instance, when  $p = 1/32$ ,  $S_{\text{Green},p}$  ended with reinforcement on one forced-choice trial, and possibly on one additional free-choice trials. Note that when  $p = 12/32$ , for each 40 Informative trials, animals had 12 reinforcers following the  $S_{\text{Green},p}$  stimulus, and 8 reinforcers following  $S_{\text{Red},1.0}$ , so the overall probability of reinforcement was 50% in both the Informative and Noninformative options. The birds were divided into four groups of two pigeons each, and each group was assigned randomly to a distinct  $p$  value. All birds received 15 sessions.

## Results and Discussion

Figure 2 shows the average preference for the Informative Option for each group across sessions. The acquisition data reveals that pigeons acquired a strong preference for the Informative Option for all probabilities of reinforcement in  $S_{\text{Green},p}$ , choosing exclusively this option during the last three sessions. This result can be explained by the obvious fact that increasing  $p$  increased the overall reinforcement rate of the Informative Option. However, the fact that steady-state preference did not decrease with  $p$  does not necessarily mean that  $p$  did not reduce the value of the Informative Option; it may mean only that for the tested range, the Informative Option remained more valuable than its alternative.

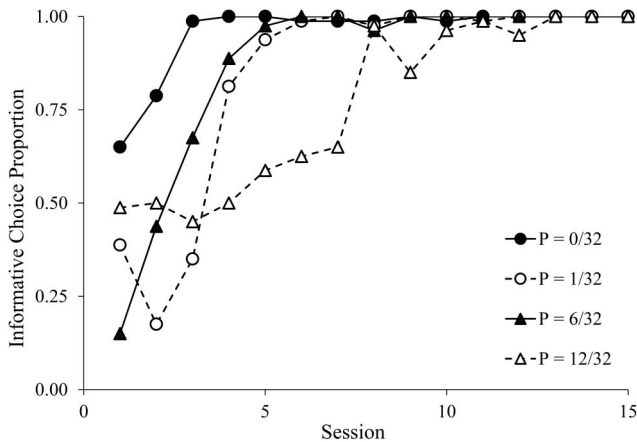


Figure 2. Mean proportion of choices of the Informative Option across sessions for each group in Experiment 1.

Even though a visual inspection of Figure 2 suggests that there were differences in the rate of preference acquisition as a function of the probability of reinforcement in  $S_{\text{Green},p}$ , a mixed analysis of variance (ANOVA) with session (15 levels) as the within-subjects factor and group (4 levels) as the between-subjects factor returned inconclusive results: a significant effect of session,  $F(14, 56) = 13.36$ ,  $p < .001$ ,  $\eta^2 = .64$ , confirmed the significant increase in preference for the Informative Option across sessions, but the effect of group was not significant,  $F(3, 4) = 0.63$ ,  $p = .64$ ,  $\eta^2 = .18$ , and the interaction was only marginally significant,  $F(42, 56) = 1.52$ ,  $p = .071$ ,  $\eta^2 = .37$ .

Regarding the rate of preference acquisition, results suggest a slowing down with  $p$  but the evidence remains weak. Nevertheless, if the value of  $p$  does affect the rate of acquisition, two interpretations are possible. On the one hand, as  $p$  increased, the probability of reinforcement in the Informative Option approached the probability of reinforcement in the Noninformative Option: .20, .225, .35, and .50 versus .50 in the Noninformative Option. If the overall probability of reinforcement affects behavior, the seemingly slower acquisition for higher values of  $p$  may be because of the increased difficulty to discriminate the probabilities of reinforcement for each option (see Bailey & Mazur, 1990, and Machado, Keen, & Macaux, 2008, for similar findings). On the other hand, the apparent decrease in rate of acquisition may mean that, as  $p$  increases, the subjective value of the Informative Option paradoxically decreases; thus, approaching the value of the alternative.

Although not conclusive, the results of Experiment 1 suggest the possibility that, as  $p$  increases the value of the Informative Options decreases. The ceiling effect observed on preference may have precluded the expression of differences in the value of the Informative Option among groups. Experiment 2 was designed to test empirically whether the value of a preferred option increases or decreases when the option is more often rewarded. To that end, we used a more sensitive measure of the value of the Informative Option: we varied the terminal-link delays of the Noninformative Option while keeping track of preference.

## Experiment 2: Probability of Food on the Informative Option With Reduced Delays on the Noninformative Option

Because all groups in Experiment 1 chose the Informative Option exclusively, a ceiling effect may have hidden differences in the Informative Option's value for different values of  $p$ . An alternative method to assess the value of the Informative Option is to increase the value of the Noninformative Option and see at which point subjects are indifferent between the two alternatives (Mazur, 1987). Thus, in Experiment 2, we systematically decreased and then increased the terminal-link delays in the Noninformative Option (similarly to Zentall & Stagner, 2011b). As the delays decrease, the Noninformative Option should become more attractive, causing the initial preference for the Informative Option to reverse. Subsequently, as the delays increase, the Noninformative Option should lose value and preference shift again to the Informative Option.

Moreover, this decrease and subsequent increase in preference for the Informative Option should be a function of  $p$ . On the one hand, if increasing  $p$  increases the value of the Informative Option, then the larger the value of  $p$ , the longer it should take for the preference function to decrease (i.e., preference for the Informative Option should decrease only with large decreases in the original 20-s Noninformative terminal-link delays) and the sooner it should subsequently increase (i.e., it should increase with only small increases in the delay from 0 s). On the other hand, if increasing  $p$  decreases the value of the Informative Option, then the larger the value of  $p$ , the sooner the preference function should decrease and the later it should subsequently increase.

From the obtained preference functions, we can also infer the indifference delays (the Noninformative terminal-link delays at which the animals are indifferent between the two options) which can be interpreted as a measure of value of the Informative Option. If, as suggested by Experiment 1, larger values of  $p$  reduce the value of the Informative Option, larger values of  $p$  should translate into longer indifference delays.

## Method

**Subjects and apparatus.** The subjects and apparatus were the same as in Experiment 1.

**Procedure.** The general procedure was the same as in Experiment 1, with the following differences. First, each session started with 80 randomly intermixed forced-choice trials (40 with each option), and ended with 80 free-choice trials, for a total of 160 trials per session. Second, all terminal-link durations started at 20 s instead of 10 s. Finally, the terminal-link durations of the Noninformative Option decreased and then increased systematically. The delays first decreased from 20 to 0 s in steps of 4 s and then increased from 0 to 20 s, again in 4-s steps. Each duration was in effect for 4 sessions, for a total of 44 sessions. Experiment 2 started immediately after Experiment 1.

## Results and Discussion

Figure 3 shows the average proportion of choices of the Informative Option as a function of the terminal-link durations in the Noninformative Option. Each curve represents a group with a

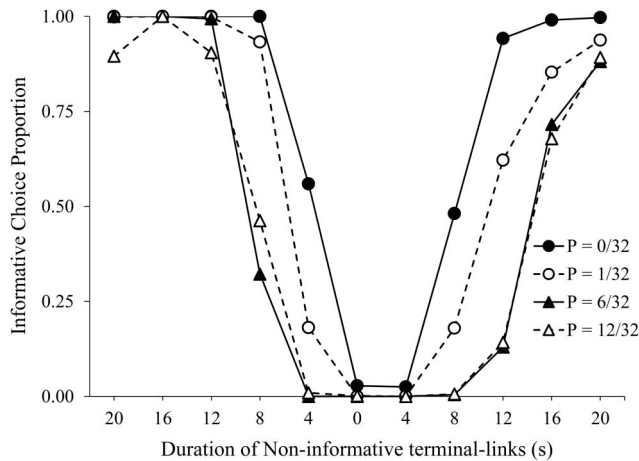


Figure 3. Mean proportion of choices of the Informative Option as a function of the noninformative terminal-link durations. Each data point averages two birds in four sessions.

different probability of reinforcement following  $S_{\text{Green},p}$ , and each data point is the average of four sessions.

First, consistent with previous findings (e.g., Fortes et al., 2017, Experiment 2; Dunn & Spetch, 1990), the leftmost data points show that all birds continued to prefer the Informative Option when all the terminal links increased from 10 (Experiment 1) to 20 s (Experiment 2). Second, as the duration of the terminal links in the Noninformative Option decreased, all birds changed their preference to the Noninformative Option; and as the duration increased, preference for the Informative Option reemerged. Moreover, the choice functions decreased and increased in an orderly fashion as predicted: As the Noninformative terminal-link delays decreased from 20 s to 0 s, preference for the Informative Option decreased first for groups with a higher probability of reinforcement following  $S_{\text{Green},p}$  ( $p = 6/32$  and  $p = 12/32$ ; see triangles in Figure 2), then for group  $p = 1/32$  (white dots), and finally for group  $p = 0/32$  (black dots). Conversely, as the Noninformative terminal-link delays increased from 0 s to 20 s, preference for the Informative Option increased, first for group  $p = 0/32$ , then for group  $p = 1/32$ , and finally for groups with a higher probability of reinforcement following  $S_{\text{Green},p}$  ( $p = 6/32$  and  $p = 12/32$ ). This symmetry in the choice functions is confirmed by the positive correlation between the indifference points calculated from the ascending and the descending data,  $r = .87$ ,  $p = .003$ . Although the curves are roughly symmetrical, they are slightly shifted to the right (i.e., the axis of symmetry is not at 0 but at 2, approximately), suggesting some carry-over effects produced by the previous duration of the Noninformative terminal-links. This result is consistent with those from Zentall and Stagner (2011b).

To estimate the indifference delay for each group, we averaged the descending and ascending indifference points (each calculated by linear interpolation from individual data). The results are presented in Figure 4; each dot is an individual subject and the line connects the average for each group. A one-way ANOVA revealed a significant effect of the probability of reinforcement on the indifference points,  $F(3, 4) = 16.92$ ,  $p = .010$ ,  $\eta_p^2 = .93$ , with post hoc Scheffé tests revealing significant differences between group

$p = 0/32$  ( $M = 6.0$  s) and groups  $p = 6/32$  ( $M = 11.8$  s,  $p = .019$ ) and  $p = 12/32$  ( $M = 12.0$  s,  $p = .017$ ).

When  $p = 0/32$ , the Noninformative Option provided 2.5 times more reinforcement than the Informative Option and pigeons still preferred the latter; animals reversed their preference only when the terminal-link duration of the Noninformative Option was reduced by 70% (from 20 to 6 s). At the other extreme, for group  $p = 12/32$ , both options yielded food on half of the trials, but the noninformative terminal-link durations had to be reduced by 41% (from 20 to 12 s) for the animals to reverse their preference. Put differently, even though the Informative Option yielded more food for group  $p = 12/32$  than for group  $p = 0/32$ , the former group required a smaller decrease in the noninformative terminal-link duration for preference to reverse. This result suggests that the value the Informative Option decreased with  $p$ .

Zentall and Stagner (2011b) also manipulated the noninformative terminal-link duration. In their experiment, the informative terminal links were 10-s long and  $p = 0$  at all times. They found that animals were indifferent between the two alternatives when the Noninformative terminal-link duration was 4.4-s long, a reduction of 56%. In the present study with  $p = 0/32$ , the delay had to be reduced by 70%, suggesting a stronger preference for the Informative Option than in Zentall and Stagner's study. In fact, the indifference delays are consistent with the preference data: Whereas in the present study (Experiment 1 and in the first sessions of Experiment 2) all birds chose the Informative Option on almost 100% of the trials, in Zentall and Stagner's experiment, choice stabilized around 80%.

## General Discussion

Animals value more an option that provides signals perfectly correlated with food (the Informative Option) than an option that provides signals uncorrelated with food (Noninformative Option), even when the latter yields more food (e.g., Mazur, 1995; Stagner & Zentall, 2010; Vasconcelos et al., 2015). The main goal of the present experiments was to assess how the value of the Informative

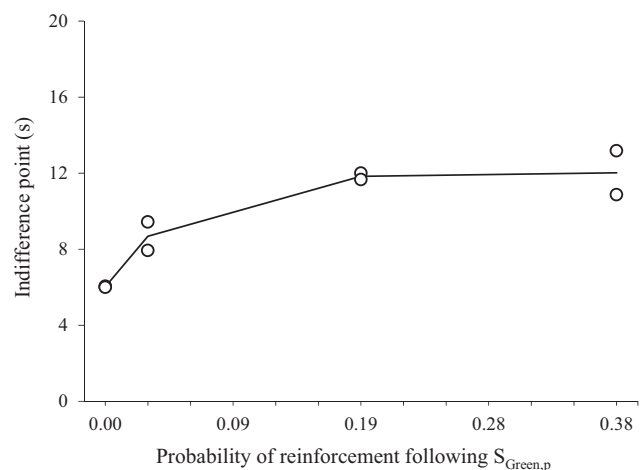


Figure 4. Indifference points (in seconds) as a function of the probability of reinforcement following the  $S_{\text{Green},p}$  stimulus. The white circles show the indifference points of individual birds and the black, connected circles show the mean indifference points.

Option changed as its probability of reinforcement increased. Our premise was that the aforementioned suboptimal preference occurs because, once the task is learned, animals do not pay attention to, or engage with, the stimulus that predicts the absence of reinforcement,  $S_{\text{Green},0,0}$  (Fortes et al., 2016; Mazur, 1995, 1996; Spetch et al., 1994; Stagner et al., 2012; Vasconcelos et al., 2015; Zentall et al., 2015). By not engaging with such a stimulus, the Informative Option becomes akin to an option that always yields food and, therefore, an option with a higher perceived value than the other partially reinforced option. By increasing the probability of reinforcement following the until-then ignored stimulus ( $S_{\text{Green},p}$  in Figure 1), we expected that pigeons would start engaging with it (Beierholm & Dayan, 2010), and the question of interest was how this manipulation would affect the value of the Informative Option.

In Experiment 1, four groups of birds were exposed to different probabilities of reward in the Informative Option. At steady state, all groups preferred exclusively the Informative Option, showing that the Informative Option remained more valued than the Noninformative Option. However, the value of the Informative Option could still have been modulated by  $p$ . We tested this hypothesis in Experiment 2 using an adjusting-delay procedure, and paradoxically found evidence that the more the  $S_{\text{Green},p}$  stimulus is reinforced (and the higher the reinforcement rate on the Informative Option), the *less value* this option has.

## Theoretical Implications

Could existing models of suboptimal choice account for these findings? To answer this question, one needs first to specify the *engagement function*,  $f$ , that relates the probability of reinforcement following a stimulus,  $p$ , to the probability of engaging that stimulus,  $f(p)$ . We have already assumed the boundary conditions  $f(0) = 0$  (the animal never attends to  $S_{\text{Green},0,0}$ ), and  $f(1) = 1$  (the animal always attends to  $S_{\text{Red},1,0}$ ), but the general form of  $f(p)$  remains unspecified.

We examined two candidates for  $f(p)$ , an all-or-none threshold function and a linear function. To predict choice, besides the functional form of  $f(p)$ , we also need to know how the costs and benefits of engaging two or more terminal-link stimuli (e.g., Red and Green keylights) combine to determine the value of an initial-link stimulus (e.g., the Informative Option). Models of choice are our best guesses as to how the combination comes about: They propose specific ways in which the costs and benefits associated with the terminal-link stimuli combine to determine the value of the initial-link stimuli, and how the latter determines behavior.

Functional models such as the RRM (Fortes et al., 2017; Fortes et al., 2016; Vasconcelos et al., 2015) stress how natural selection acts on the long-term fitness consequences of foraging activities; they focus on ultimate causes. Mechanistic models such as the HDM (Mazur, 1984, 1987, 1997) stress the psychological processes driving behavior; they focus on proximate causes.

Briefly, the RRM is based on optimal foraging principles and was developed by Vasconcelos and colleagues (2015) to account for the suboptimal preference found in the task depicted in Figure 1. According to this model, when given a choice, animals choose the option that maximizes the rate of energy intake, computed as the ratio of the expected energy gain from food to the expected time spent to obtain that food. However, in this computation not all durations are considered. When a stimulus is never followed by

food, as it is the case for the  $S_{\text{Green},p}$  when  $p = 0$ , this stimulus is not attended to. Similarly to the RRM, the HDM assumes that any stimulus not followed by reinforcement ( $S_{\text{Green},0,0}$ ) does not become a conditioned reinforcer and is “ignored” by animals. The question of interest then is whether the models can also capture the modulation of the value of the Informative Option when  $S_{\text{Green}}$  is occasionally reinforced (i.e., when  $p > 0$  in  $S_{\text{Green},p}$ ).

Appendixes A and B review the RRM and the HDM, respectively, and derive their value functions with an all-or-none and a linear engagement function. The all-or-none engagement function assumes that if a stimulus is followed by reinforcement with probability greater than a threshold, the animal always engages the stimulus; otherwise, the animal never engages the stimulus. The linear engagement function, on the other hand, assumes that the likelihood of engaging a stimulus is determined by the probability of reinforcement after that stimulus: the higher the probability, the more likely the animal is to engage.

Figure 5 shows the results. The left and right panels show the results from the RRM and the HDM, respectively. The top and bottom panels plot the value of the Noninformative (dashed) and Informative (continuous) options when the model includes an all-or-none and a linear engagement function, respectively. At a qualitative level, the HDM and RRM make similar predictions (cf. left vs. right panels). If the engagement function is all-or-none, the animals should prefer the Informative Option when  $p$  is lower than a threshold  $\theta$ , and switch to the Noninformative Option when  $p > \theta$  (until  $p = .375$ ); if the engagement function is linear, even though the difference in value between the two options decreases and then increases again, animals should always prefer the Informative Option regardless of  $p$ .

The main difference between the all-or-none and linear engagement functions is that the first predicts a preference reversal, whereas the second does not. However, exclusive preference for the Informative Option independently of  $S_{\text{Green},p}$ 's reinforcement probability does not suffice to conclude that the engagement function is linear rather than all-or-none. It is also necessary to show that, even though the Informative Option always had more value than the Noninformative Option, its value actually decreased with  $p$ .

In Experiment 1, all groups preferred exclusively the Informative Option—a result more parsimoniously accommodated by a linear than an all-or-none engagement function. However, if the function is indeed linear, the value of the Informative Option ought to be negatively modulated by  $p$  (cf. lower left panel of Figure 5). In fact, the results of Experiment 2 are consistent with the hypothesis that the more a stimulus is reinforced, the more the animals engage with it, thus, supporting a linear engagement hypothesis. To account for the results of Experiment 1, an all-or-none engagement function would require a threshold  $\theta > .375$ , but under no circumstance would this function predict the modulation of value by  $p$  observed in Experiment 2. Although other monotonically increasing functions predict the same pattern of results, the main implication of our findings is that the engagement function does not seem to depend on a threshold but vary gradually with  $p$ .

Figure 6 shows, for each value of  $p$ , the average indifference delays obtained in Experiment 2 (black dots), as well as the indifference delays predicted by the RRM (white dots) and the HDM (white triangles), both with a linear engagement function. Although both models predict the observed trend (the correlation

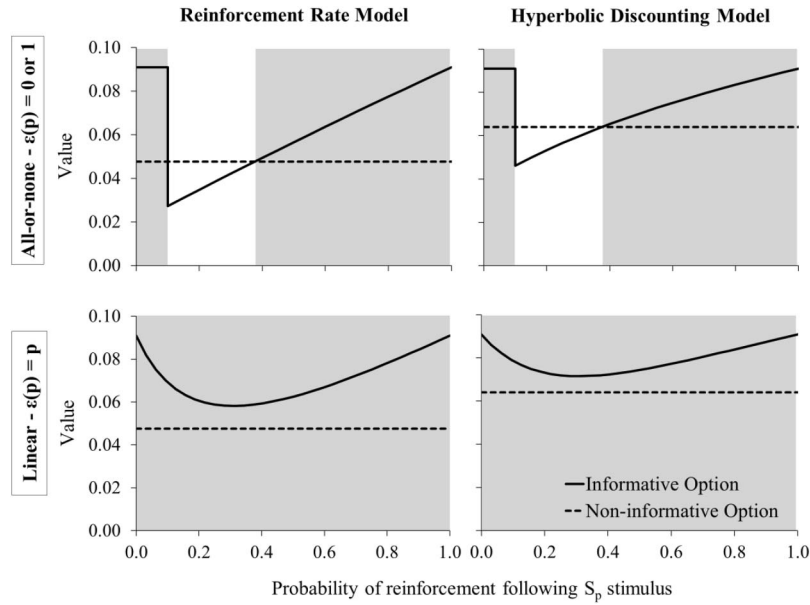


Figure 5. Predicted values of the Informative and Noninformative options as a function of the probability of reinforcement following  $S_{\text{Green},p}$ . The left panels show the predictions of the Reinforcement Rate Model with an all-or-none function (top) and a linear engagement function (bottom); the right panels show the predictions of the Hyperbolic Discounting Model also with an all-or-none function (top) and a linear engagement function (bottom). A value of  $D = 10$  s was used in all predictions, and for the Hyperbolic Discounting Model,  $K$  was set to 1. Other values for the parameters influence both the Informative and the Noninformative options, so the quantitative predictions may differ but the qualitative ones remain the same. The shaded areas indicate the parameter space where a preference for the Informative Option is predicted.

between observed and predicted values was  $r = .97$ ,  $p = .026$  for HDM, and  $r = .97$ ,  $p = .028$ , for RRM), they overestimate the indifference points. Other parameter values may bring the models' predictions closer to the data but because our main interest was to

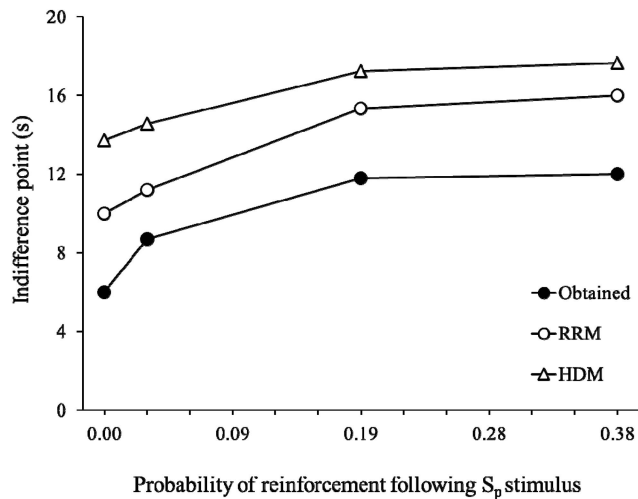


Figure 6. Indifference points observed in Experiment 2 and predicted by the HDM (Hyperbolic Discounting Model) and the RRM (Reinforcement Rate Model) as a function of the probability of reinforcement following the  $S_{\text{Green},p}$  stimulus.  $D = 20$  s for the informative terminal-link delay;  $K = 1$  for the HDM.

examine the engagement function and show how functional and mechanistic models can go hand-in-hand, we did not pursue the model fitting issues further.

An alternative account for our findings may be provided by Information Theory (Coombs, Dawes, & Tversky, 1970; Shannon & Weaver, 1949). Broadly, Information Theory predicts that the higher the reward uncertainty associated with a stimulus, the less value it has. Thus, a stimulus that signals a probability of reinforcement of 0 or 1 has minimum uncertainty and, therefore, maximum value, whereas a stimulus that signals a probability of reinforcement of .5 has maximum uncertainty and minimum value (Shahan & Cunningham, 2015). In the present task, as the probability of reinforcement following  $S_{\text{Green},p}$  increased, the predictive value of  $S_{\text{Red},1.0}$  remained constant, but the uncertainty associated with  $S_{\text{Green},p}$  increased, which in turn should decrease preference for the Informative Option. This prediction is similar to the prediction of the HDM and the RRM with a linear engagement function, with the exception that these two models predict a minimum value at a probability around .31 (cf. Figures 2 and 6), whereas Information Theory predicts a minimum value at a probability around .50 (at the value of maximum uncertainty). Note, however, that this account demands (a) the specification of the mechanism(s) by which the uncertainty of the terminal links translates into attractiveness or value of the initial link, and (b) the interaction between uncertainty and reinforcement. Assume, for instance, that both terminal links of a given option have minimum uncertainty because they both predict the *absence* of reinforcement. The attractiveness of the initial link should nonetheless be

very low to none even though both terminal links convey maximum information. At the moment, we do not advance a proposal for how this may take place.

Be that as it may, the Information Theory account seems at odds with other findings in the suboptimal choice literature. For example, Roper and Zentall (1999) gave pigeons a choice between two options with the same probability of reinforcement. As usual, the Informative Option provided discriminative stimuli for food and the Noninformative Option provided nondiscriminative stimuli. Across conditions, the overall probability of reinforcement was 0.125, 0.5, or 0.875 but it remained the same for the two options. As the probability of food increased, the Informative Option should maintain a high value because the terminal-link stimuli continued to convey maximum information. In contrast, the uncertainty associated with the Noninformative Option increased up to  $P_{\text{food Noninfo}} = .5$  and then decreased (i.e., uncertainty follows an inverted U-shape centered at  $p = .5$ ). Thus, according to the Information Theory hypothesis, preference for the Informative Option should follow an inverted U-shape as  $P_{\text{food Noninfo}}$  ranges from 0 to 1, with a maximum at  $P_{\text{food Noninfo}} = .5$ . Roper and Zentall (1999) found that, consistent with Information Theory, preference for the Informative Option slightly decreased as the probability increased from 50 to 87.5%. However, inconsistent with Information Theory, preference for the Informative Option slightly increased as the probability decreased from 50 to 87.5%.

In another study, Zentall and Stagner (2011a) found that animals chose suboptimally even when there was no uncertainty associated with the options' outcomes. In their experiment, the Informative Option comprised a stimulus followed by 10 pellets or a different stimulus followed by the absence of food (0 pellets); the Noninformative Option comprised a stimulus followed always by 3 pellets. Although the information conveyed by all stimuli was maximal, pigeons preferred the Informative Option that sometimes provided a higher amount of food. These findings are hard to reconcile with the Information Theory hypothesis.

In conclusion, the results of the present experiments are yet another demonstration of suboptimal choice, with the additional counterintuitive finding that the more an option is reinforced, the less value it may have. When the probability of reinforcement following  $S_{\text{Green}}$  was zero, the Informative Option had more value than when the probability was greater than zero, suggesting that with  $p = 0$  the delays to no food are not taken into account, but as  $p$  increases some of those delays contribute to the perceived rate. Also, we showed that the results are consistent with the higher the probability of food, the higher the probability of engagement. We only tested probabilities up to .375. If the engagement function is indeed linear, we should expect an increase in value for the Informative Option for higher probabilities of reinforcement. Future research should test a wider range of probabilities and directly contrast alternative models.

## References

- Bailey, J. T., & Mazur, J. E. (1990). Choice behavior in transition: Development of preference for the higher probability of reinforcement. *Journal of the Experimental Analysis of Behavior*, 53, 409–422. <http://dx.doi.org/10.1901/jeab.1990.53-409>
- Bateson, M., & Kacelnik, A. (1996). Rate currencies and the foraging starling: The fallacy of the averages revisited. *Behavioral Ecology*, 7, 341–352. <http://dx.doi.org/10.1093/beheco/7.3.341>
- Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation. *Journal of the Experimental Analysis of Behavior*, 12, 861–874. <http://dx.doi.org/10.1901/jeab.1969.12-861>
- Beierholm, U. R., & Dayan, P. (2010). Pavlovian-instrumental interaction in 'observing behavior'. *PLoS Computational Biology*, 6, e1000903. <http://dx.doi.org/10.1371/journal.pcbi.1000903>
- Cardinal, R. N., & Aitken, M. R. (2010). Whisker: A client-server high-performance multimedia research control system. *Behavior Research Methods*, 42, 1059–1071. <http://dx.doi.org/10.3758/BRM.42.4.1059>
- Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction*. Englewood Cliffs, NJ: Prentice Hall, Inc.
- Dunn, R., & Spetch, M. L. (1990). Choice with uncertain outcomes: Conditioned reinforcement effects. *Journal of the Experimental Analysis of Behavior*, 53, 201–218. <http://dx.doi.org/10.1901/jeab.1990.53-201>
- Fortes, I., Machado, A., & Vasconcelos, M. (2017). Do pigeons (*Columba livia*) use information about the absence of food appropriately? A further look into suboptimal choice. *Journal of Comparative Psychology*, 131, 277–289. <http://dx.doi.org/10.1037/com0000079>
- Fortes, I., Vasconcelos, M., & Machado, A. (2016). Testing the boundaries of "paradoxical" predictions: Pigeons do disregard bad news. *Journal of Experimental Psychology: Animal Learning and Cognition*, 42, 336–346. <http://dx.doi.org/10.1037/xan0000114>
- Machado, A., Keen, R., & Macaux, E. (2008). Making analogies work: A selectionist model of choice. In N. K. Innis (Ed.), *Reflections on adaptive behavior: Essays in honor of J. E. R. Staddon* (pp. 23–50). Cambridge, MA: The MIT Press.
- Mazur, J. E. (1984). Tests of an equivalence rule for fixed and variable reinforcer delays. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 426–436. <http://dx.doi.org/10.1037/0097-7403.10.4.426>
- Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Volume 5. Effects of delay and of intervening events on reinforcement value* (pp. 55–73). Hillsdale, NJ: Erlbaum.
- Mazur, J. E. (1989). Theories of probabilistic reinforcement. *Journal of the Experimental Analysis of Behavior*, 51, 87–99. <http://dx.doi.org/10.1901/jeab.1989.51-87>
- Mazur, J. E. (1995). Conditioned reinforcement and choice with delayed and uncertain primary reinforcers. *Journal of the Experimental Analysis of Behavior*, 63, 139–150. <http://dx.doi.org/10.1901/jeab.1995.63-139>
- Mazur, J. E. (1996). Choice with certain and uncertain reinforcers in an adjusting-delay procedure. *Journal of the Experimental Analysis of Behavior*, 66, 63–73. <http://dx.doi.org/10.1901/jeab.1996.66-63>
- Mazur, J. E. (1997). Choice, delay, probability, and conditioned reinforcement. *Animal Learning & Behavior*, 25, 131–147. <http://dx.doi.org/10.3758/BF03199051>
- Mazur, J. E. (2005). Effects of reinforcer probability, delay, and response requirements on the choices of rats and pigeons: Possible species differences. *Journal of the Experimental Analysis of Behavior*, 83, 263–279. <http://dx.doi.org/10.1901/jeab.2005.69-04>
- Mazur, J. E., & Ratti, T. A. (1991). Choice behavior in transition: Development of preference in a free-operant procedure. *Animal Learning & Behavior*, 19, 241–248. <http://dx.doi.org/10.3758/BF03197882>
- Pisklak, J. M., McDevitt, M. A., Dunn, R. M., & Spetch, M. L. (2015). When good pigeons make bad decisions: Choice with probabilistic delays and outcomes. *Journal of the Experimental Analysis of Behavior*, 104, 241–251. <http://dx.doi.org/10.1002/jeab.177>
- Roper, K. L., & Zentall, T. R. (1999). Observing behavior in pigeons: The effect of reinforcement probability and response cost using a symmetrical choice procedure. *Learning and Motivation*, 30, 201–220. <http://dx.doi.org/10.1006/lmot.1999.1030>

- Shahan, T. A., & Cunningham, P. (2015). Conditioned reinforcement and information theory reconsidered. *Journal of the Experimental Analysis of Behavior*, 103, 405–418. <http://dx.doi.org/10.1002/jeab.142>
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Spetch, M. L., Mondloch, M. V., Belke, T. W., & Dunn, R. (1994). Determinants of pigeons' choice between certain and probabilistic outcomes. *Animal Learning & Behavior*, 22, 239–251. <http://dx.doi.org/10.3758/BF03209832>
- Stagner, J. P., Laude, J. R., & Zentall, T. R. (2012). Pigeons prefer discriminative stimuli independently of the overall probability of reinforcement and of the number of presentations of the conditioned reinforcer. *Journal of Experimental Psychology: Animal Behavior Processes*, 38, 446–452. <http://dx.doi.org/10.1037/a0030321>
- Stagner, J. P., & Zentall, T. R. (2010). Suboptimal choice behavior by pigeons. *Psychonomic Bulletin & Review*, 17, 412–416. <http://dx.doi.org/10.3758/PBR.17.3.412>
- Stephens, D. W., & Anderson, D. (2001). The adaptive value of preference for immediacy: When shortsighted rules have farsighted consequences. *Behavioral Ecology*, 12, 330–339. <http://dx.doi.org/10.1093/beheco/12.3.330>
- Vasconcelos, M., Monteiro, T., & Kacelnik, A. (2015). Irrational choice and the value of information. *Scientific Reports*, 5, 13874. <http://dx.doi.org/10.1038/srep13874>
- Zentall, T. R. (2016). Resolving the paradox of suboptimal choice. *Journal of Experimental Psychology: Animal Learning and Cognition*, 42, 1–14. <http://dx.doi.org/10.1037/xan0000085>
- Zentall, T. R., Laude, J. R., Stagner, J. P., & Smith, A. P. (2015). Suboptimal choice by pigeons: Evidence that the value of the conditioned reinforcer rather than its frequency determines choice. *The Psychological Record*, 65, 223–229. <http://dx.doi.org/10.1007/s40732-015-0119-2>
- Zentall, T. R., & Stagner, J. (2011a). Maladaptive choice behaviour by pigeons: An animal analogue and possible mechanism for gambling (sub-optimal human decision-making behaviour). *Proceedings Biological Sciences*, 278, 1203–1208. <http://dx.doi.org/10.1098/rspb.2010.1607>
- Zentall, T. R., & Stagner, J. P. (2011b). Sub-optimal choice by pigeons: Failure to support the Allais paradox. *Learning and Motivation*, 42, 245–254. <http://dx.doi.org/10.1016/j.lmot.2011.03.002>

## Appendix A

### The Reinforcement Rate Model (RRM)

Consider a foraging situation where a predator searches for a prey item of unit energy ( $E = 1$ ) during time  $s$ ; upon encountering the prey it pursues it for time  $D$ ; if it captures the prey (probability  $P$ ) it handles it during time  $h$ ; if it does not capture it (probability  $1 - P$ ) it searches again. This would yield a rate of

$$R = \frac{P}{s + P \times (D + h) + (1 - P) \times D} = \frac{P}{s + D + P \times h} \quad (\text{A1})$$

Applying Equation A1 to the present task,  $P$  is the probability of reinforcement in an option,  $s$  is the duration of the intertrial interval (ITI),  $D$  is the duration of the signals for food, and  $h$  is the time of access to food.

In its present form, however, Equation A1 ignores the informational imbalance of the task: Whereas the terminal links of the Informative Option reduce the uncertainty about whether or not a reward is due—they are informative, the terminal links of the Noninformative Option do not reduce reward uncertainty—they convey no information. In its final form, the RRM takes into account the usability of the information conveyed by the terminal links of the Informative Option: The information that a stimulus is always followed by food can be used differently than the information that a stimulus is never followed by food. To that end, Vasconcelos et al. (2015) made two additional assumptions.

First, when  $p = 0$ , as when the typical  $S_{\text{Green},0,0}$  is shown, this stimulus is not attended to. The reason for the assumption comes from the natural foraging behavior of animals in the wild. While searching, if an animal encounters a cue indicating unambiguously that a prey item is not available, the animal will most likely stop searching it and will search for another prey elsewhere. For this reason, when the animal is presented with  $S_{\text{Green},0,0}$  in the Informative Option, its durations are not included in the rate calculation. Second, because research has shown that only events that precede the reward are taken into account (e.g., Stephens & Anderson, 2001), the ITI that follows the rewards also is not taken into account in the rate computation (i.e.,  $s$  is removed from Equation A1; for a more detailed explanation see Vasconcelos et al., 2015; see also Bateson & Kacelnik, 1996; Fortes et al., 2016; Mazur, 1989). Thus, assuming  $h = 1$ , the perceived rate on the Informative Option is given by:

$$\begin{aligned} R_{\text{Info}} &= \frac{P}{s + P_{\text{Red}} \times (D + 1) + P_{\text{Green}} \times D} = \frac{P}{P_{\text{Red}} \times (D + 1)} \\ &= \frac{1}{1 + D} \end{aligned} \quad (\text{A2})$$

Note that for this option  $p = 1$ , because all events related to Green are ignored and the Informative Option becomes functionally similar to an option that is always reinforced.

(Appendices continue)

For the Noninformative Option, the durations  $D$  and  $h$  are always taken into account because all terminal-link stimuli are associated with reinforcement. Using Equation A1 and considering that  $P_{Blue} + P_{Yellow} = 1$ , the perceived reinforcement rate (with  $h = 1$ ) for this option is:

$$R_{Non-info} = \frac{0.5}{.5 + 0.2 \times [0.5 \times (D + 1) + 0.5 \times D] + 0.8 \times [0.5 \times (D + 1) + 0.5 \times D]} = \frac{1}{1 + 2D} \quad (A3)$$

That is, provided that  $D > 0$ , the RRM predicts that  $R_{Info} > R_{Noninfo}$  and, therefore, that animals should prefer the Informative Option. Yet, if some reinforcement follows the Green stimulus (i.e.,  $S_{Green,0.0}$  is replaced by  $S_{Green,p}$ , with  $p > 0$ ), the animal may attend to Green and one of the assumptions used to derive  $R_{Info}$  (i.e., that animals ignore the Green stimulus, see Equation A2) is not warranted. To determine the consequences of attending to Green, we need to define the engagement function and include it in the RRM.

### All-or-None Engagement Function

According to an all-or-none engagement function, if a stimulus is followed by reinforcement with probability greater than a threshold  $\theta$ , the animal always engages the stimulus; if the stimulus is followed by reinforcement with probability less than  $\theta$ , the animal never engages the stimulus. Formally,  $f(p)$  is defined as follows

$$f(p) = \begin{cases} 0, & p \leq \theta \\ 1, & p > \theta \end{cases} \quad (A4)$$

with  $0 < \theta < 1$ . The function satisfies the boundary conditions mentioned above,  $f(0) = 0$  and  $f(1) = 1$ .

Given the step function with  $h = 1$  and  $D = 10$ , let us assume the threshold for engagement equals  $\theta = 0.1$ . Depending on the value of  $p$ , the animal may or may not engage with  $S_{Green,p}$  in the Informative Option. For  $p \leq \theta$ , the animal engages only with  $S_{Red,1.0}$ . Thus, to receive one unit of food the animal spends  $D + 1$  units of time (we assume  $h = 1$ ). For  $p > \theta$ , the animal engages both with  $S_{Red,1.0}$  and with  $S_{Green,p}$ , but whereas the former is always followed by reinforcement, the latter is followed by reinforcement only on a proportion  $p$  of these trials. The reinforcement rate of the Informative Option is then given by

$$R_{Info}(p) = \begin{cases} \frac{.2 + .8 \times \frac{1}{1+D}}{.2 \times (D+1) + .8 \times [p \times (D+1) + (1-p) \times D]} & p \leq \theta \\ \frac{1+4p}{1+4p+5D} & p > \theta. \end{cases} \quad (A5)$$

Regarding the Noninformative Option, the animal engages on all trials ( $f(p) = 1$ ), and, therefore, the animal receives, on average, half a reinforcer on each trial. Hence, it has to spend about  $D + 0.5$  s to receive half a reinforcer. The reinforcement rate is

$$R_{Non-info} = \frac{1}{1+2D} \quad (A6)$$

To illustrate, assume  $\theta = .1$ —the animal engages any stimulus followed by reinforcement on more than 10% of the trials. The top left panel of Figure 5 shows the rates of the Noninformative (dashed line) and Informative (continuous line) options as a function of  $p$ .

The rate of the Noninformative Option is independent of  $p$  (see Equation A6). The rate of the Informative Option depends on how  $p$  compares with  $\theta$ : While  $p \leq \theta$ , the animal does not engage the  $S_{Green,p}$  stimulus, the reinforcers following that stimulus are not taken into account, and the rate of the Informative Option remains unchanged and greater than the rate of the Noninformative Option. For  $p > \theta$ , the animal always engages the stimulus and the rate of the Informative Option increases with  $p$  from 0.027 to 0.091. For  $.1 \leq p \leq .375$ , the Informative Option has less value than the Noninformative Option; for  $p > .375$ , the Informative Option has more value than the Noninformative Option. Hence, if engagement follows a step function, the animals should display three preference regimens as  $p$  varies from 0 to 1: Initially, they should prefer the Informative Option, then reverse preference, and then reverse preference once again. The first reversal occurs at the threshold,  $\theta$ , and the second at the value of  $p$  that equalizes the overall reinforcement probability in the two options (.375 in the example).

### Linear Engagement Function

In a linear engagement function, the probability of reinforcement after a stimulus determines the likelihood of engagement: the higher the probability, the more likely the animal is to engage. The only linear function satisfying the two boundary conditions,  $f(0) = 0$  and  $f(1) = 1$ , is

$$f(p) = p \quad (A7)$$

Given this engagement function the reinforcement rate in the Informative Option Equals

(Appendices continue)

$$R_{Info}(p) = \frac{.2 + .8p^2}{.2(D+1) + .8p[p(D+1) + (1-p)D]} = \frac{1}{1 + \frac{1+4p}{1+4p^2}D} \quad (A8)$$

In the Noninformative Option, because the probability of reinforcement is .5 the animal only engages on half of the trials and, thus, only takes into account half of the reinforcers and the delays to obtain them. The rate of food intake in the Noninformative Option is given by

$$R_{Non-info} = \frac{.5 \times .5}{.5 \times .5 \times (D+1) + .5 \times .5 \times D} = \frac{1}{1+2D} \quad (A9)$$

Thus, the reinforcement rate in the Noninformative Option remains the same when the engagement function is all-or-none or linear.

The bottom left panel of Figure 5 plots the two functions. The value of the Noninformative Option remains constant, but the value of the Informative Option changes nonlinearly with  $p$ . As  $p$

increases, and, as a consequence, as the overall probability of reinforcement in this option increases, its value follows a U-shaped function, with minimum at  $p \approx .31$ . For all  $p$ , the Informative Option has more value than the Noninformative Option, thus, predicting a preference for the former.

In summary, with a linear engagement function we expect the Informative Option to be preferred regardless of the probability of reinforcement in  $S_{Green,p}$ . If the decision rule is winner-take-all, the animals should prefer the Informative Option exclusively. However, if the decision rule is less extreme—perhaps choice depends in a smooth way on the ratio or the difference between the two option values (e.g., Baum & Rachlin, 1969)—then the proportion of choices of the Informative Option should be above .5, but have a shape similar to the value of the Informative Option in Figure 5. That is, a stronger preference for the Informative Option at the lowest and highest values of  $p$ , but approaching indifference for intermediate values of  $p$ .

## Appendix B

### The Hyperbolic Discounting Model (HDM)

According to the HDM (Mazur, 1984, 1987), the value of a delayed reward is given by

$$V = \frac{A}{1+KD} \quad (B1)$$

where  $V$  is the value of the reward,  $A$  is related to the amount of the reward,  $D$  is the delay to the reward, and  $K$  is a discounting rate parameter that determines how rapidly value decreases with delay. In most cases, setting  $K = 1$  provides good estimates of value for pigeons (Mazur, 2005), so henceforth this value will be used. Moreover, because we did not vary reward amount and are concerned mostly with ordinal predictions, we also let  $A = 1$ .

Equation B1 applies only when a reinforcer is delivered after a fixed delay,  $D$ , since choice because with probabilistic rewards the delay between choice and reward varies.

#### The HDM With Probabilistic Reinforcement

According to the HDM, a probabilistic reinforcer is functionally equivalent to a set of reinforcers delivered after different delays, with the constraint that these delays include only the time spent in the presence of stimuli at least partially reinforced (e.g., Mazur, 1989; for empirical tests of these rules, see Mazur, 1995, 1996; Mazur & Ratti, 1991). Put differently, stimuli never followed by reinforcement are disregarded and excluded from the calculations.

In the target task, when the animal chooses one option, it is presented with one of two stimuli, A or B, say, with probabilities  $r_A$  and  $r_B$ , respectively, with  $r_A + r_B = 1$ . When stimulus A is presented, one of three mutually exclusive events happens: event  $A^+$ , the animal engages with the stimulus and after the delay it receives a reward; event  $A^-$ , the animal engages with the stimulus, but after the delay it does not receive a reward; and event  $A^0$ , the animal does not engage with the stimulus, in which case the stimulus duration and the trial outcome are simply ignored. An obvious extension to stimulus B defines the events  $B^+$ ,  $B^-$ , and  $B^0$ .

Let  $p_A$  denote the probability of reinforcement at the end of stimulus A, and  $f(p_A)$  the probability the animal engages with stimulus A when it occurs; the equivalent probabilities for stimulus B are  $p_B$  and  $f(p_B)$ , respectively. Then, as Figure B1 shows, the probabilities of the three events involving stimulus A and the three events involving stimulus B are

$$\begin{aligned} P(A^+) &= r_A \cdot f(p_A) \cdot p_A & ; & \quad P(B^+) = r_B \cdot f(p_B) \cdot p_B & ; \\ P(A^-) &= r_A \cdot f(p_A) \cdot (1 - p_A) & ; & \quad P(B^-) = r_B \cdot f(p_B) \cdot (1 - p_B) & ; \\ P(A^0) &= r_A \cdot [1 - f(p_A)] & ; & \quad P(B^0) = r_B \cdot [1 - f(p_B)] & . \end{aligned} \quad (B2)$$

As a check, note that the sum of all probabilities equals  $r_A + r_B = 1$ , as it should.

(Appendices continue)

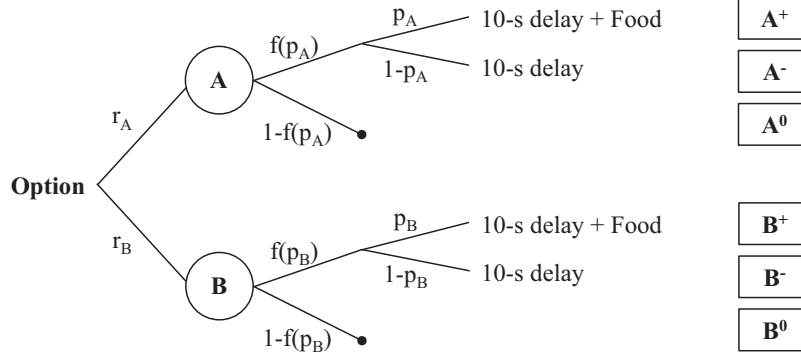


Figure B1. Probability tree for the possible events in the Informative and Noninformative options. Each possible trial type was labeled according to whether the animal engaged and was reinforced (+), engaged but was not reinforced (−), or did not engage with the stimulus (0).  $r_A$  denotes the probability of encountering stimulus A,  $p_A$  the probability of reinforcement at the end of stimulus A, and  $f(p_A)$  the probability the animal engages with stimulus A when it occurs. The same notations are used for stimulus B.

When the animal does not engage with a stimulus, it is as if the corresponding trial had not taken place. These nonengaged trials do not affect an option's value and may be discarded. Their probability equals

$$\begin{aligned} P(A^0 \text{ or } B^0) &= r_A \cdot [1 - f(p_A)] + r_B \cdot [1 - f(p_B)] \\ &= r_A + r_B - r_A \cdot f(p_A) - r_B \cdot f(p_B) \\ &= 1 - r_A \cdot f(p_A) - r_B \cdot f(p_B), \end{aligned} \quad (B3)$$

and, therefore, the probability of the remaining trials, those that affect value, equals

$$\begin{aligned} P(A^+ \text{ or } A^- \text{ or } B^+ \text{ or } B^-) &= \overline{P(A^0 \text{ or } B^0)} \\ &= 1 - P(A^0 \text{ or } B^0) \\ &= r_A \cdot f(p_A) + r_B \cdot f(p_B). \end{aligned} \quad (B4)$$

To discard nonengaged trials we need to renormalize the probabilities of the events  $A^+$ ,  $A^-$ ,  $B^+$ , and  $B^-$  so that they add to 1. But first, to simplify the notation, we define the weight function  $w$ ,

$$w(r_A, p_A, r_B, p_B) = \frac{r_A \cdot f(p_A)}{r_A \cdot f(p_A) + r_B \cdot f(p_B)}. \quad (B5)$$

Then, the renormalization yields the new probabilities,<sup>1</sup>

$$\begin{aligned} P(A^+) &= w(r_A, p_A, r_B, p_B) p_A & ; \\ P(A^-) &= w(r_A, p_A, r_B, p_B) (1 - p_A) & ; \\ P(B^+) &= w(r_B, p_B, r_A, p_A) p_B & ; \\ P(B^-) &= w(r_B, p_B, r_A, p_A) (1 - p_B) & . \end{aligned} \quad (B6)$$

Again, we note that their sum equals  $w(r_A, p_A, r_B, p_B) + w(r_B, p_B, r_A, p_A) = 1$ .

To compute the value of the option we need to determine the probability that reinforcement will occur at the end of a D-s delay. In terms of events, reinforcement means that  $A^+$  or  $B^+$  occurred and, therefore, its probability, denoted  $P(+)$ , equals  $P(A^+) + P(B^+)$ . That is,  $P(+)$  is a weighted average of  $p_A$  and  $p_B$ ,

$$P(+) = w(r_A, p_A, r_B, p_B) p_A + w(r_B, p_B, r_A, p_A) p_B, \quad (B7)$$

That simplifies to

$$P(+) = p_B + (p_A - p_B) w(r_A, p_A, r_B, p_B). \quad (B8)$$

Given that reinforcement occurs with probability  $P(+)$  at the end of each trial, and that successive trials are independent, the distribution of the number of trials before a reinforcer occurs is Geometric with parameter  $P(+)$ . If we let  $P(i)$  denote the probability a reinforcer will occur at the end of  $i$  trials ( $i = 1, 2, \dots$ ), then

$$P(i) = P(+) \times [1 - P(+)]^{i-1}. \quad (B9)$$

Finally, a reinforcer after  $i$  trials corresponds to a cumulative delay of  $i \times D$  seconds. Hence, for the HDM the value of the option is given by

$$V = \sum_{i=1}^{\infty} P(i) \frac{A}{1 + K \cdot i \cdot D}. \quad (B10)$$

<sup>1</sup> To be more rigorous we should have written  $P(A^+ | A^0 \text{ or } B^0)$  for the probability of event  $A^+$  given that we discard nonengaged trials, and similarly for  $A^-$ ,  $B^+$ , and  $B^-$ . Henceforth, we assume the analysis is restricted to engaged trials.

To summarize, to compute the value of an option that with probabilities  $r_A$  and  $r_B = 1 - r_A$  presents stimuli A and B, and after a D-s delay presents a reward with probabilities  $p_A$  and  $p_B$ , respectively, we use equations B5, B8, B9, and B10 repeated here for convenience:

$$\begin{aligned} w(r_A, p_A, r_B, p_B) &= \frac{r_A \cdot f(p_A)}{r_A \cdot f(p_A) + r_B \cdot f(p_B)} \\ P(+) &= p_B + (p_A - p_B)w(r_A, p_A, r_B, p_B) \\ P(i) &= P(+) \times [1 - P(+)]^{i-1} \\ V &= \sum_{i=1}^{\infty} P(i) \frac{A}{1 + K \cdot i \cdot D}. \end{aligned} \quad (B11)$$

### The Original Suboptimal Choice Task

In the original task, for the Informative Option, A and B correspond to the Red and Green stimuli, respectively, and therefore,  $r_A = 0.2$ ,  $r_B = 0.8$ ,  $p_A = 1$ ,  $p_B = 0$ , and  $D = 10$  s. Although the engagement function remains undefined, we have already assumed two boundary conditions,  $f(0) = 0$  (i.e., the animal does not engage a stimulus never reinforced), and  $f(1) = 1$  (i.e., the animal always engages a stimulus always reinforced). Hence, with  $A = K = 1$  and  $D = 10$  s, we get

$$\begin{aligned} w(r_A, p_A, r_B, p_B) &= \frac{0.2 \cdot f(1)}{0.2 \cdot f(1) + 0.8 \cdot f(0)} = 1 \\ P(+) &= 0 + (1 - 0) \cdot 1 = 1 \\ P(i) &= \begin{cases} 1, & i = 1 \\ 0, & i > 1 \end{cases} \\ V_{\text{Info}} &= \frac{1}{1 + 10} \approx 0.091. \end{aligned} \quad (B12)$$

For the Noninformative Option, A and B correspond to the Blue and Yellow stimuli, respectively, and therefore,  $r_A = 0.2$ ,  $r_B = 0.8$ ,  $p_A = 0.5$ ,  $p_B = 0.5$ , and  $D = 10$  s. Hence

$$\begin{aligned} w(r_A, p_A, r_B, p_B) &= \frac{0.2 \cdot f(0.5)}{0.2 \cdot f(0.5) + 0.8 \cdot f(0.5)} = 0.2 \\ P(+) &= 0.5 + (0.5 - 0.5) \cdot 0.2 = 0.5 \\ P(i) &= 0.5(1 - 0.5)^{i-1} = 0.5^i \\ V_{\text{Non-Info}} &= \sum_{i=1}^{\infty} 0.5^i \frac{1}{1 + 10 \cdot i} \approx 0.064. \end{aligned} \quad (B13)$$

The value of the Noninformative option is less than the value of the Informative option, which accounts for the nonoptimal choice in the task.

### The Modified Suboptimal Choice Task

If the Green stimulus is reinforced with probability  $p$ , but all other parameters remain unchanged (i.e.,  $p_A = f(p_A) = 1$  and

$p_B = p$ ), then the value of the Noninformative Option does not change,<sup>2</sup> but the value of the Informative Option depends on the specific form of  $f(p)$ . After some algebra we obtain

$$\begin{aligned} V_{\text{Non-Info}} &= \sum_{i=1}^{\infty} \frac{0.5^i}{1 + 10 \cdot i} \approx 0.064 \\ V_{\text{Info}}(p) &= \sum_{i=1}^{\infty} \frac{P(i)}{1 + 10 \cdot i}, \\ \text{with } P(i) &= \left( \frac{1 + 4f(p)p}{1 + 4f(p)} \right) \left[ \frac{4f(p)(1-p)}{1 + 4f(p)} \right]^{i-1}. \end{aligned} \quad (B14)$$

We consider now two specific forms for  $f(p)$ .

**All-or-none engagement function.** Suppose the animal engages stimulus S (A or B in Figure B1) if and only if the probability of reinforcement following S exceeds the threshold  $\theta$ . That is,

$$f(p) = \begin{cases} 0, & p \leq \theta \\ 1, & p > \theta \end{cases}. \quad (B15)$$

Then,

$$\begin{aligned} V_{\text{Non-Info}} &= \sum_{i=1}^{\infty} \frac{0.5^i}{1 + 10 \cdot i} \approx 0.064 \\ V_{\text{Info}}(p) &= \begin{cases} \frac{1}{1 + 10} \approx 0.091, & p \leq \theta \\ \frac{0.2 + 0.8p}{0.8(1-p)} \sum_{i=1}^{\infty} \frac{[0.8(1-p)]^i}{1 + 10 \cdot i} & p > \theta. \end{cases} \end{aligned} \quad (B16)$$

Figure 5 (top right panel) plots Equations B16 for  $\theta = 0.1$ .  $V_{\text{Info}}$  remains constant at 0.091 for  $p \leq \theta$  and then jumps to the monotonically increasing curve given by the second expression in Equation B16. When  $p = 12/32 = 0.375$  (and assuming  $\theta \leq 0.375$ ),  $V_{\text{Info}}$  equals  $V_{\text{Noninfo}}$ , for then the animal always engages and the reinforcement rates in the two options equal 0.5 (Informative Option:  $0.2 \times 1 + 0.8 \times 0.375 = 0.5$ ; Noninformative Option:  $0.2 \times 0.5 + 0.8 \times 0.5 = 0.5$ ).

**Linear engagement function.** Suppose that  $f(p) = p$ . In this case,

$$\begin{aligned} V_{\text{Non-Info}} &= \sum_{i=1}^{\infty} \frac{0.5^i}{1 + 10 \cdot i} \approx 0.064 \\ V_{\text{Info}}(p) &= \sum_{i=1}^{\infty} \frac{P(i)}{1 + 10 \cdot i}, \text{ with } P(i) = \left( \frac{1 + 4p^2}{1 + 4p} \right) \left[ \frac{4p(1-p)}{1 + 4p} \right]^{i-1}. \end{aligned} \quad (B17)$$

<sup>2</sup> If  $p_A = p_B$  in Equation B8,  $P(+)$  is not affected by the weight function  $w$  and hence by the engagement function  $f$ .

(Appendices continue)

Figure 5 (bottom right panel) plots Equations B17.  $V_{\text{Info}}$  is concave up, reaching a minimum close to  $p = .31$ , but it is always greater than  $V_{\text{Noninfo}}$ .

When  $K * D$  is large compared with 1, the expression  $(1 + K * D * i)$  is close to  $K * D * i$ . In this case we can approximate the value of an option by replacing  $(1 + K * D * i)$  by  $K * D * i$  (or, because we let  $K = 1$ , replacing  $1 + D * i$  by  $D * i$ ), and then using the following identity,

$$\sum_{i=1}^{\infty} \frac{a^i}{i} = -\ln(1 - a), \quad |a| < 1. \quad (\text{B18})$$

Therefore, for  $V_{\text{Noninfo}}$  we get

$$\begin{aligned} V_{\text{Non-Info}}^{\text{Approx}} &= \sum_{i=1}^{\infty} \frac{0.5^i}{1 + 10 \cdot i} \\ &\approx \frac{1}{10} \sum_{i=1}^{\infty} \frac{0.5^i}{i} \\ &= \frac{-\ln(0.5)}{10} \\ &\approx 0.069. \end{aligned} \quad (\text{B19})$$

and for  $V_{\text{Info}}$  we get

$$\begin{aligned} V_{\text{Info}}^{\text{Approx}}(p) &= \sum_{i=1}^{\infty} \frac{P(i)}{1 + 10 \cdot i} \\ &\approx \frac{1}{10} \sum_{i=1}^{\infty} \frac{P(i)}{i} \\ &= \frac{1}{10} \left( \frac{1 + 4p^2}{4p(1 - p)} \right) \sum_{i=1}^{\infty} \left[ \frac{4p(1 - p)}{1 + 4p^2} \right]^i / i \\ &= \frac{1}{10} \left( \frac{1 + 4p^2}{4p(1 - p)} \right) \ln \left( \frac{1 + 4p^2}{1 + 4p^2} \right). \end{aligned} \quad (\text{B20})$$

Received March 8, 2017

Revision received December 15, 2017

Accepted December 18, 2017 ■