Frequency effects in action versus value learning

Hilary J. Don & Darrell A. Worthy

Texas A&M University

Accepted on 1/25/21 – *Journal of Experimental Psychology Learning Memory and Cognition*

Correspondence should be addressed to Hilary J. Don, who is now at
Division of Psychology and Language Sciences
University College London
26 Bedford Way, Bloomsbury, London WC1H 0AP, United Kingdom
h.don@ucl.ac.uk

Abstract

Recent work in reinforcement learning has demonstrated a choice preference for an option that has a lower probability of reward (A) when paired with an alternative option that has a higher probability of reward (C), if A has been experienced more frequently than C (the *frequency-effect*). This finding is critical as it is inconsistent with widespread assumptions that expected value is based on average reward, and instead suggests that value is based on cumulative instances of reward. However, option frequency may also affect instrumental reinforcement of choosing A during training, which may then transfer to choice on AC trials. This study therefore aimed to assess the contribution of action reinforcement and option value to the frequency-effect across two experiments. In both experiments we included an additional test phase in which participants were asked to rate the likelihood of reward for each choice option, a response that should be unaffected by action reinforcement. In Experiment 1, participants completed the original choice training phase. In Experiment 2, participants were presented with each option individually, thus removing reinforcement of choice during training. Single cue training reduced the strength of the preference for A compared to choice training, suggesting a contributing role of action reinforcement. However, frequency effects were still evident in both experiments. We found that the pattern of reward likelihood ratings were consistent with the pattern of choice preferences in both experiments, suggesting that action reinforcement may also influence judgements about the likelihood of receiving reward.

Making optimal choices requires an evaluation of the expected value of each alternative option. Estimates of the likelihood of future rewarding outcomes are assumed to arise through trial-and-error experience with each option. Different reinforcement learning models make different assumptions about how expected values are estimated. For instance, popular *Delta rule* models value options according to the probability that they will provide a rewarding outcome (Rescorla & Wagner, 1972; Widrow & Hoff, 1960; Williams, 1992). Alternatively, other models value options according to the cumulative number of rewarding options they have provided in the past, such as the *Decay rule* model (e.g. Erev & Roth, 1998; Yechiam & Busemeyer, 2005; Yechiam & Ert, 2007). Although typically options that have a higher probability of reward will also provide rewards more frequently, this is not necessarily the case if the amount of experience with each choice option differs.

This distinction between reward frequency and reward probability dates back to work by Estes (1976a, 1976b). In Estes' seminal work, participants viewed a series of observational trials presenting winning and losing pairs of stimuli, for example, the results of an opinion poll. Prior to each trial of the experiment, a hypothetical individual in the hypothetical population would be asked by the computer about their opinions between two alternatives, such as two political candidates, or two health care products (simply referred to as stimuli in Estes's studies). Participants could then view the preference of the hypothetical individual. Participants were told to observe the results across the series of trials and attempt to form a mental impression of the relative likelihoods that different stimuli would be preferred by the individuals being sampled. Participants were then presented with different combinations of stimuli and asked to predict which would be the likely preferred alternative in a sample from the previously surveyed population. These studies reliably found greater predictions that the stimuli that had been presented frequently would be preferred over stimuli that had been presented less frequently, even if the alternative had a higher probability of winning. Estes' work suggests that judgments are made based on memory for the frequency of events, rather than probability per se (see also Brainerd, 1981; Einhorn & Hogarth, 1981).

In a recent study (Don, Otto, Cornwall, Davis & Worthy, 2019), we found similar preferences in a reinforcement learning task using a similar trial structure to Estes, where participants learned to choose between different option pairs to receive reward. On some trials, they selected between A (.65 reward probability) and B (.35 reward probability), where A had a higher probability of reward. On other trials, they selected between C (.75 reward probability) and D (.25 reward probability), where C had a higher probability of reward, and also the highest probability of reward of all four options. Critically however, there were twice as many AB trials than CD trials. This means that although C had the highest probability of reward, A will have provided a greater number of rewards throughout the task. Participants then completed a transfer phase where they chose between different combinations of options. The critical test was on AC transfer trials, which paired the option with a higher probability of reward (C) with the option that had provided a greater number of rewards (A). If people value options based on the probability of reward, they should prefer option C, but if they value options based on the cumulative frequency of reward, they should prefer option A. We found that participants showed a consistent preference for option A, indicating a *frequency effect*.

This finding, combined with Estes's prior work demonstrating frequency effects in observational learning (1976a; 1976b), calls into question the assumption that people learn and value options based on reward probability, and instead suggests that value is more likely to be based on the cumulative number of rewards provided by each option in the past. These findings also relate to classic studies that suggest frequency information is automatically coded (Ekstrand, Wallace & Underwood, 1966; Hintzman, 1988). This has important implications for reinforcement learning models, and our understanding of the factors that drive learning and decision making, particularly as many prominent reinforcement learning models value options according to average reward, such as those using Delta updating rules (e.g. Rescorla & Wagner, 1972; Widrow & Hoff, 1960; Williams, 1992). Indeed, models using a Delta rule to update expected value were unable to account for this effect (Don et al., 2019). Instead, the choice effect was better anticipated and fit by models using a Decay rule to update expected values, which increments expected value each time an option is rewarded (Erev & Roth, 1998; Yechiam & Busemeyer, 2005; Yechiam & Ert, 2007).

However, decision making is also influenced by factors beyond expected value. Researchers have suggested that decision making is separable into two components (Barto, 1992; 1995; O'Doherty et al., 2007). The first is learning the expected value of each option through experience. This value learning can emerge through simple contingent pairings of a stimulus and reward. The second is referred to as action selection, which is learning the action of choosing the option that provides the greatest reward. This is essentially the product of instrumental reinforcement. That is, the more an action is reinforced, the greater the likelihood of repeating that action in the future (Sutton & Barto, 1998; Thorndike, 1911).

Action selection is assumed to be based on estimates of option value (Barto, 1992; 1995). Nevertheless it is possible that option frequency could affect both option value and action selection components separately. For instance, in the previously described task, the frequency of reward provided by option A may increase its value according to cumulative updating of expected value, like the process assumed by the Decay rule. However, as there are a greater number of AB trials than CD trials, the action of choosing A will be reinforced more frequently than the action of choosing C, assuming participants learn the optimal choices on these trials. That is, the action of "choosing A over the alternative" will have more instances of reinforcement than "choosing C over the alternative". Thus, the more strongly reinforced action of choosing A may be more likely to be repeated when participants encounter AC trials at test. This kind of instrumental conditioning may therefore bias participants to choose A, even if the learning rule that determines option value does not actually favour option A over option C.

Estes (1976a, 1976b) demonstrated that frequency effects occurred when learning about preferred alternatives were purely observational. These findings might suggest that action selection plays little role in these effects, as they occur when no choices are made. However, in a reinforcement learning task, participants are actively making choices to receive points, which may be more reinforcing than passively observing the winner of an opinion poll, as in Estes' studies. Thus, although similar results were found in an observational task, it does not prohibit the possibility that action selection is a contributing factor to the strength of the effect in our paradigm. The current study therefore aimed to further test the involvement of action selection to frequency effects, using the same reinforcement learning task as Don et al. (2019) described above. First, we included an additional test phase for which action tendencies should be irrelevant. In this phase, each option was presented individually, and participants were required to rate the likelihood of receiving reward, given the presented option was chosen. As there are no binary choices involved, action selection should not influence these ratings. Thus, any difference in ratings for each option may better reflect their expected value. Second, we tested whether removing choice between alternative options from training, and therefore removing greater reinforcement of choosing A over the alternative than choosing C over the alternative, effectively removes or reduces any effects caused by option frequency differences at test. To do this, we designed a task that retained the majority of task elements from Experiment 1, while removing reinforcement of choice between two options. For this reason, we decided against using an observational version of the task, where the computer chooses between the two options on each trial, and participants observe the resulting points. Although this would remove choice reinforcement, we wanted participants to be actively involved in receiving rewards for their actions during training, as in Experiment 1. The critical element of the task that may contribute to the A-preference is greater reinforcement of the action of choosing A when two options are presented than choosing C when two options are presented, as this action is most likely to transfer to AC test trials where two options are presented. Experiment 2 therefore presented each option individually during training. Participants were asked to "pass" or "play" each option, and choosing to play the option provided the same probability of reinforcement as those in Experiment 1. Note that this design does not eliminate all forms of instrumental reinforcement from the task, as participants will still receive differing amounts of reinforcement for "playing" each option. However, this is a different instrumental response to the alternative forced choice that is likely to generalise to AC test trials. The important thing is that in this training condition, participants will not have more experience and reinforcement of "choosing A over the alternative" than "choosing C over the alternative" that could transfer to test trials. Throughout the remainder of the paper, we refer to action reinforcement as reinforcement of this specific alternative choice response. Third, we tested whether Delta and Decay models could account for differences in these training conditions (choice vs. no choice) without appealing to an independent action selection mechanism.

In Don et al. (2019), we assessed frequency effects by comparing choice to chance. While this indicates whether there is a significant preference for A over C, it does not provide a direct test of the influence of option frequency on choice. To provide a better test of the influence of option frequency and reward probability on choice preferences, we introduced two comparison groups in which we manipulate either the difference in reward probability provided by A and C options, or trial frequency of A and C options during training. We will refer to the original design as the *probability x frequency group*, as it manipulates both the probability of reward associated with each option, and the frequency with which each option is presented. In one comparison group, each option had the same probability of reward as the probability x frequency group, such that the probability of reward was higher for C than A, but each pair of options was presented in equal base-rates. We will refer to this group as the *probability-only* group, as only the probabilities associated with each option differ, but the frequency of presentation of A and C do not. Comparing performance in the probability x frequency condition to the probability only condition indicates the effect of option frequency on choice and ratings. Note that this comparison gives us an indication of the effect of *option* frequency on choice (i.e. the difference in base-rates between A and C) as opposed to the effect of *reward* frequency per se. Within the probability only group, we cannot distinguish between the effect of reward probability and reward frequency on choice, as C provides both a higher probability of reward and also a higher frequency of reward when base-rates are equal. However, we can gauge how participants make choices in this task when there is no difference in option base-rates. In the other comparison group, A and C had equal probability of reward, but there were twice as many A trials than C trials. We will refer to this group as the *frequency-only* group, as the reward probabilities of A and C do not differ, but their base-rates do. Thus, in this group, any differences in preference for A and C at test should be a result of frequency differences alone. Comparing the probability x frequency group to the frequency only group will show whether there is any effect of differences in reward probability on choice and ratings. It is worth noting that Estes (1976b, Study 5) found a strong effect of frequency in a similar situation where two pairs had equal reward probability, but one was presented more frequently, thus we predict a strong effect of frequency here as well.

To summarise, we ran two experiments to test the possibility that the strength of the frequency-effect is influenced by more frequent reinforcement of choosing A over the alternative, independent of differences in expected value. Experiment 1 replicated the choice task from Don et al. (2019), with the addition of two comparison groups, and a ratings test phase. Experiment 2 trained cues individually in each of the three groups, also with the additional ratings test phase. Finally, we simulated and fit the data from these two training conditions with Delta and Decay models, to determine whether any differences in choice preferences can be accounted for by expected value alone. If action reinforcement is contributing to the frequency-effect, we should expect different results across choice and ratings test phases. Additionally, we should expect effects of option frequency on choice to be removed or reduced in Experiment 2.

Experiment 1

Experiment 1 aimed to replicate the effect shown in Don et al. (2019), and introduced two comparison conditions, and a ratings test phase. The design of the task is shown in Table 1. We will refer to the original design as the *probability x frequency* group, which has twice as many AB than CD trials, and C has the highest probability of reward. The original task used the following probabilities of reinforcement: A = .65, B = .35, C = .75 and D =.25. The difference in probability between .65 and .75 may be difficult to discern, and so we adjusted the probabilities to A = .65, B = .35, C = .80 and D = .20. In this way, the probability of A reinforcement is 15% above chance, and the probability of C reinforcement is 15% above that of A. The probability-only comparison group maintains the probability differences in the probability x frequency group, but trains AB and CD in equal frequency. This condition allows us to assess how participants respond to the probabilities of rewards in the absence of base-rate differences, and also to assess the effect of frequency on choice by comparison with the probability x frequency group. The frequency-only group matches the base-rate differences of AB and CD trials in the probability x frequency group, but both A and C have equal probability of reward. In this condition, the reward probabilities were A = .70, B = .30, C = .70 and D = .30. We chose these values as one option is clearly more optimal than the other, and .70 is also a value somewhere between the reinforcement rate of A and C. This allows us to assess the effect of frequency on choice when there are no differences in reward probability, as well as the influence of probability differences on choice when compared with the probability x frequency group. We expect a preference for A over C on AC trials in the probability x frequency and frequency only groups. We also expect more A choices in the probability x frequency group than the probability only group if the frequency of A influences choice preferences. We will differentiate between a preference for A (choice of A greater than chance), and an effect of option frequency (the difference in responses between the probability x frequency and probability only groups), as it is possible to have an effect of option frequency without a significant preference for A. We predicted that the ratings phase would be less affected by an instrumentally reinforced tendency to select one alternative over another.

Method

Participants

The experiment received ethical approval from the Institutional Review Board (IRB) at Texas A&M University (IRB2019-0663D). Based on the previous sample sizes used in Don et al. (2019), we aimed to recruit a sample size of at least 35 participants per group. One-hundred and twenty participants from Texas A&M University participated in return for partial course credit, and were randomly allocated to each group. To ensure choices in the transfer phase were based on good learning of the reward contingencies during training, we introduced an inclusion criterion of at least 50% optimal choices during training. Eleven participants did not pass this criterion and were excluded from the analyses (four in the probability x frequency group, four in the probability only group, and three in the frequency only group). Of the remaining 109 participants, 80 were female (mean age = 18.5, *SD* = 0.85).

Table 1.

Task design

		Option			
Group		A	B	C	D
Probability x frequency	p(reward)	.65	.35	.80	.20
	Base-rate	2	2	1	1
Probability only	p(reward)	.65	.35	.80	.20
	Base-rate	1	1	1	1
Frequency only	p(reward)	.70	.30	.70	.30
	Base-rate	2	2	1	1

Stimuli and apparatus

The experiment was programmed using PsychToolbox 3 for Matlab (Kleiner, Brainard, & Pelli, 2007), and was run on PC computers in groups of up to 5 participants. Cue stimuli were four 200 x 300 pixel rectangles representing different decks of cards, presented horizontally aligned on the screen (see Figure 1). For each participant, each option A-D was randomly allocated to a card colour and position. Each option remained in the same position throughout the entire task, with the exception of the rating phase, where each cue was presented individually in the horizontal center of the screen.

Procedure

Training. In the training phase, participants were instructed to choose from decks of cards in order to gain points. They were informed that their goal was to gain as many points as possible, and to learn which decks were the most rewarding. On each trial, two cues were presented on the screen, accompanied by a prompt to “pick a card”. Participants made their choice by clicking on a card. Once a card was chosen, the selected card was “flipped” and turned white, and the points won were displayed on the card. If the trial was a reward trial, “+10” was shown in green text, and if it was a no-reward trial, “0” was shown in black text. The feedback remained on screen for 1500 ms, followed by a 500 ms inter-trial interval, before presenting the next pair of cards. A point tally was present on the bottom of the screen throughout training. There were four blocks of 30 trials. In the probability x frequency and frequency only groups, AB trials were presented 20 times per block, and CD trials were presented 10 times per block. In the probability only group, both AB and CD trials were presented 15 times each.

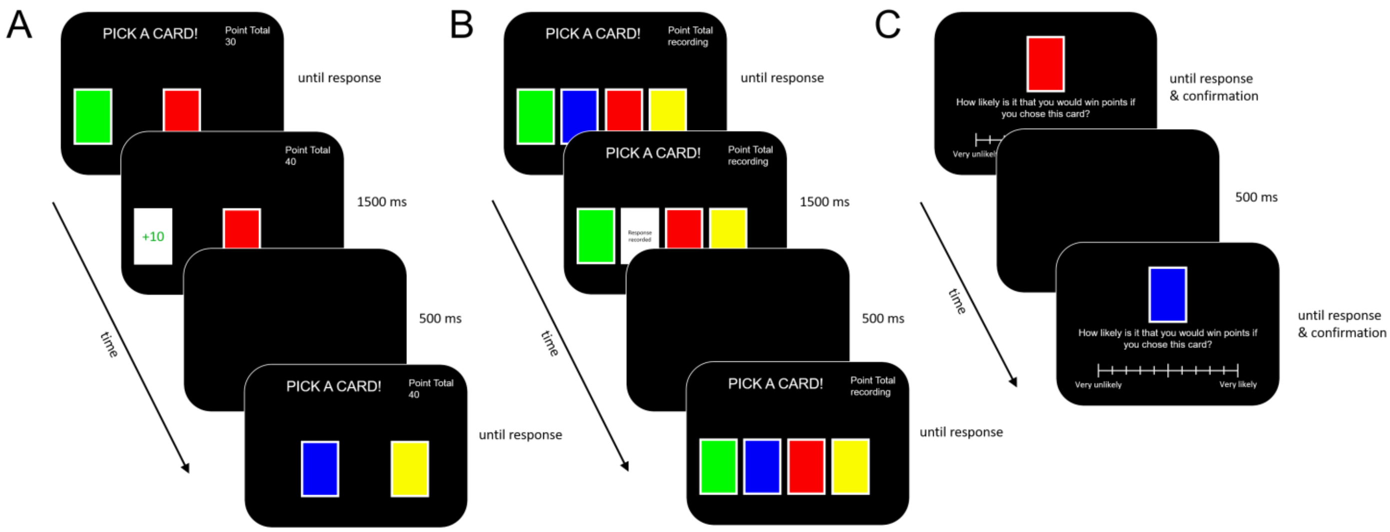

Figure 1. Schematic of the reward task. A) During training, two options were presented on each trial. Once an option was selected it turned white, and the amount of points earned was displayed on the card during training. During the transfer phase the card turned white and the card instead displayed “response recorded”. There was an inter-trial interval of 500 ms between trials. B) In the four-choice phase, participants chose between all four options. C) In the rating phase, each deck was presented individually and participants rated how likely it was that they would receive reward.

Transfer test. In the transfer test, participants were instructed to choose between different combinations of cards. They were told that they would continue to earn points for their choices, but would not see how many until the end of the experiment. On each trial, one of the combinations of test trials shown in Table 1 were presented. When a card was selected, the card turned white and “response recorded” was displayed on the card. The point tally displayed “Points: recording” There were 10 blocks of the transfer phase, with each trial type presented twice per block.

Four-choice test. In the next test phase, participants were able to choose between all four options. All possible cards were displayed on each trial, and participants were not provided feedback. Responses were displayed in the same manner as the previous phase. For brevity, and because the results closely follow those for the critical AC test trials, the data from this test phase are presented in the Supplementary Material.

Likelihood ratings. Participants were asked to rate how likely it was that they would win points if they had chosen each deck of cards. On each trial, one of the cards was displayed in the center of the screen. Participants made their rating on an 11-point linear analogue scale that ranged from “Very unlikely” to “Very likely”. Participants were able to adjust their rating before pressing the space bar to continue (without feedback). At the end of the experiment, participants were shown a tally of the points they had earned throughout the entire task.

Results & Discussion

For the critical analyses, we included p-values as well as Bayes factors to assess the strength of evidence for the alternative hypothesis (BF_{10}). Bayes factors were computed in JASP using Bayesian ANOVAs or t-tests with the default priors. Typically, a Bayes factor between 1 and 3 indicates minimal support, between 3 and 10 indicates moderate support, and greater than 10 indicates strong support for the alternative hypothesis. However, they can also be interpreted continuously as the odds in favour of the alternative hypothesis (Wagenmakers et al., 2018). For Bayesian ANOVA, Bayes factors for the main effects indicate the likelihood of the data given the main effects model relative to a null model (BF_{10}). Bayes Factors on interaction effects indicate evidence for the interaction by comparing models including the interaction effect with models excluding the effect (BF_{incl}; Rouder et al., 2017).

Training

The mean proportion of optimal choices on AB (A choices) and CD (C choices) across training for each group are presented in Figure 2. We analysed the data comparing the probability x frequency group with each of the comparison groups in two separate 2 x 2 x 5 mixed measures ANOVAs, with group as a between-subjects factor, and trial type (AB vs. CD) and block (1-5) as within-subjects factors.

Probability x frequency vs. probability only. Comparing the probability x frequency and probability only groups, there was a significant main effect of group $F(1,69) = 10.21$, $p = .002$, $\eta_p^2 = .129$, $BF_{10} = 13.77$, indicating greater overall optimal choices in the probability only group. There was also a main effect of trial type, indicating a greater proportion of optimal choices on CD trials than AB trials, $F(1,69) = 26.54$, $p < .001$, $\eta_p^2 = .278$, $BF_{10} = 1.17 \times 10^{11}$, which did not interact with group, $F(1,69) = 0.86$, $p = .356$, $\eta_p^2 = .012$, $BF_{incl} = 0.32$. There was a significant linear effect of block, indicating an increase in optimal choices across the course of training, $F(1,69) = 31.18$, $p < .001$, $\eta_p^2 = .311$, and this did not interact with group, $F(1,69) = 1.15$, $p = .288$, $\eta_p^2 = .016$, or trial type, $F(1,69) = 0.03$, $p = .858$, $\eta_p^2 < .001$. Thus while the probability only group responded more optimally overall, there were no differences in relative learning on CD and AB trials between groups.

Probability x frequency vs. frequency only. Comparing the probability x frequency and frequency only groups, there was no significant main effect of group, $F(1,71) = 0.004$, $p = .952$, $\eta_p^2 < .001$, $BF_{10} = 0.30$. There was also no significant main effect of trial type, $F(1,71) = 0.021$, $p = .886$, $\eta_p^2 < .001$, $BF_{10} = 0.26$, but trial type interacted with group, $F(1,71) = 13.46$, $p < .001$, $\eta_p^2 = .159$, $BF_{incl} = 3.0$, such that there was a higher proportion of optimal choices for CD than AB trials in the probability x frequency group, but a higher proportion of optimal choices for AB than CD in the frequency only group. There was also a significant linear effect of block, $F(1,71) = 32.62$, $p < .001$, $\eta_p^2 = .315$, and this interacted with trial type, $F(1,71) = 5.24$, $p = .025$, $\eta_p^2 = .069$, but not group, $F(1,71) = 1.69$, $p = .198$, $\eta_p^2 = .023$. There was a significant three-way interaction between the linear effect of block, trial type, and group, $F(1.71) = 9.54$, $p = .003$, $\eta_p^2 = .118$. This indicates a greater rate of learning on AB trials than CD trials that was more evident in the frequency only group than the probability x frequency group.

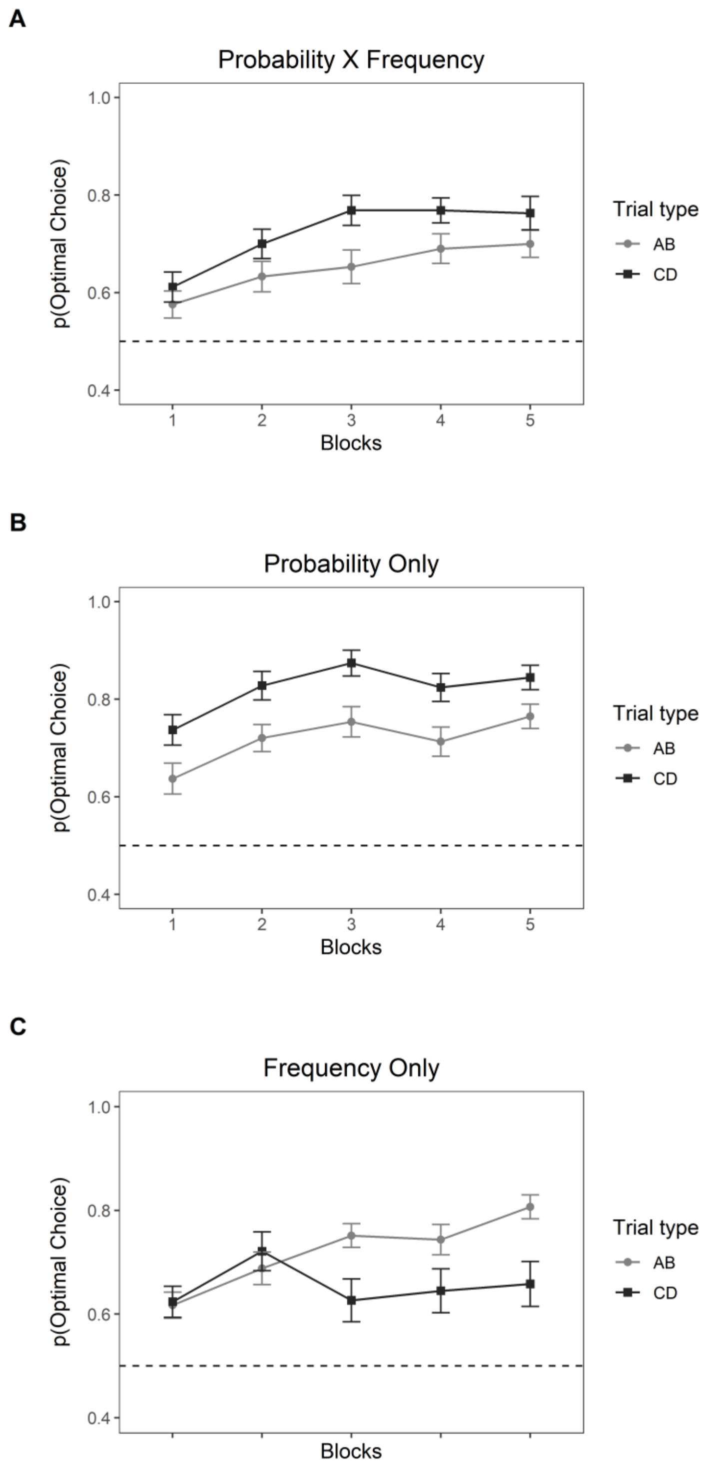

Figure 2. Proportion of optimal choices during training for a) probability x frequency group, b) probability group, c) frequency group in Experiment 1

AC test trials

The proportion of C choices on AC trials in each group is shown in Figure 3. Responding above chance (.50) indicates a preference for option C, while responding below chance indicates a preference for option A. In the probability x frequency group, the proportion of C choices were significantly below chance, indicating a preference for option A (*M* = .40, *SEM* = .05), $t(34) = -2.21$, $p = .034$, $BF_{10} = 1.55$. Therefore, this experiment replicated the preference for A on AC trials when A had a lower probability of the outcome than C, but had been presented more frequently than C. In the probability only group, where AB and CD trials were experienced in equal base-rates, participants responded more optimally, with C choices significantly above chance (*M* = .70, *SEM* = .05), $t(35) = 4.14$, $p < .001$, $BF_{10} = 128.54$. In the frequency only group, where both A and C had equal probability but AB trials were experienced more frequently, there was also a significant preference for the more frequent option A (*M* = .31, *SEM* = .04), $t(37) = -4.75$, $p < .001$, $BF_{10} = 729.68$. The probability x frequency group had significantly fewer C choices than the probability only group, $t(69) = -4.50$, $p < .001$, $BF_{10} = 700$, indicating that option frequency had a significant effect on responding. There was no significant difference in responding in the probability x frequency and frequency only groups, $t(71) = 1.33$, $p = .186$, $BF_{10} = 0.52$.

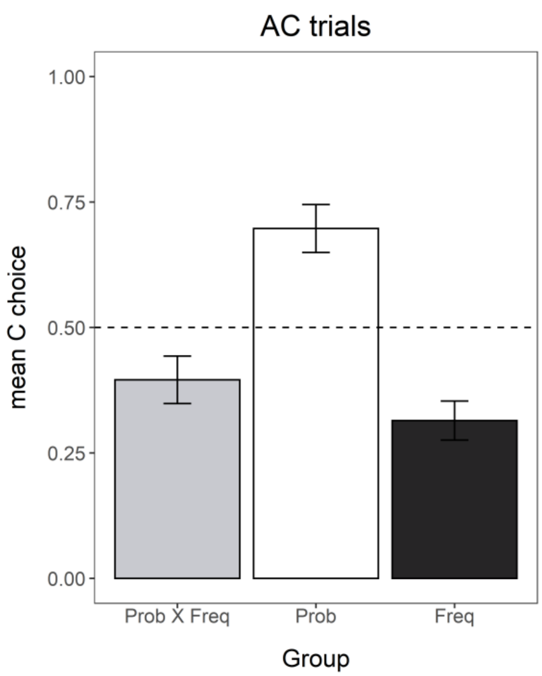

Figure 3. Proportion of C choices on AC trials in the transfer test in Experiment 1. The dashed line indicates chance level performance.

Likelihood ratings

We also assessed whether participants judgment of the likelihood of reward provided by each option was affected by frequency. Figure 4 shows mean likelihood ratings for each option. Analysis of the ratings focused primarily on the A and C options. We again compared the probability x frequency group with each of the control groups in two separate 2 x 2 mixed measures ANOVAs with group as a between-subjects factor and trial type (A vs. C) as a within-subjects factor.

Probability x frequency vs. probability only. There was no significant main effect of group, F(1,69) = 1.99, p = .163, η_p^2 = .028, BF_{10} = 0.41, or trial type, F(1,69) = 1.95, p = .168, η_p^2 = .027, BF_{10} = 0.46. However, there was a significant interaction between group and trial type, F(1,69) = 17.98, p < .001, η_p^2 = .207, BF_{incl} = 1556.65. Here, ratings were higher for C than A in the probability only group, but higher for A than C in the probability x frequency group. This indicates that the higher frequency of A led participants to judge it as more likely to provide reward than the less frequent, higher probability option C.

Probability x frequency and frequency only. There was a significant main effect of trial type, where A was rated as more effective than C, $F(1,71)$ = 11.03, p = .001, η_p^2 = .135, BF_{10} = 70.36, but no main effect of group, $F(1,71)$ = .391, p = .534, η_p^2 = .005, BF_{10} = 0.23, and no interaction, $F(1,71)$ = .263, p = .610, η_p^2 = .004, BF_{incl} = 0.27. Thus, participants' judgments of the likelihood of reward were higher for the frequent option regardless of probability of reward.

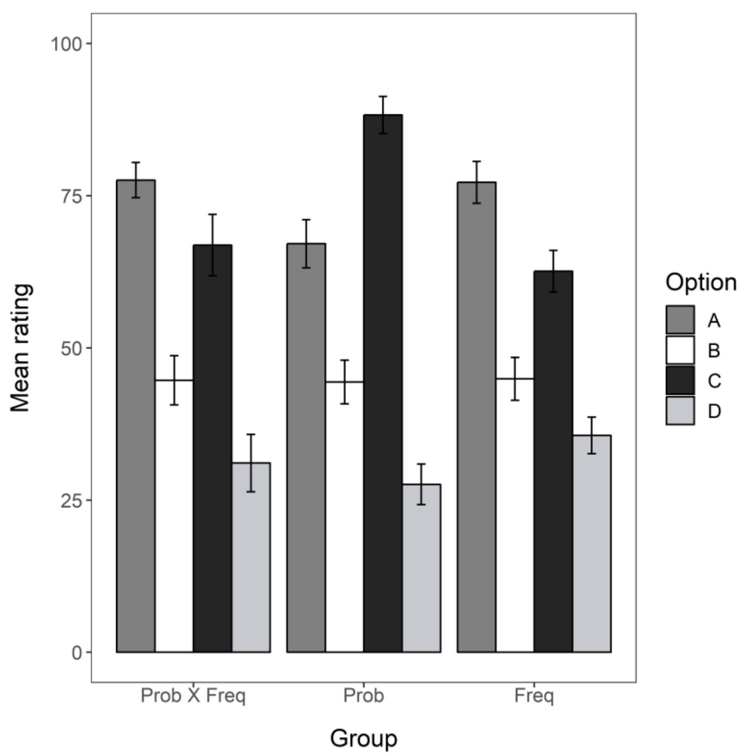

Figure 4. Mean ratings of the likelihood of receiving reward for each option in Experiment 1.

Thus, in Experiment 1, we observe a preference for A in both the probability x frequency group, and the frequency only group. We also observed significantly different responses between the probability x frequency only group, and probability only groups, indicating a clear effect of option frequency on choice. On the other hand, there were no significant differences between the probability x frequency group and the frequency only group, suggesting that preferences in these groups are driven primarily by option frequency regardless of the probability of reward associated with each option. Within the probability only group, participants were able to learn and use probabilities in a more optimal manner when all choice options were presented in an equal base-rate. Here, participants reliably chose the higher probability option, C, and rated it as more likely to provide reward. It is worth noting again that we cannot distinguish between learning reward probability and reward frequency within this group, as the higher probability option would also provide a greater number of rewards when the base-rates are equal. However, this group allows a better test of the effect of option frequency, as optimal choices and ratings for option C were reduced in the probability x frequency group compared to the probability-only group. Responding at test in the probability x frequency group is unlikely to be due to a failure to adequately learn the optimal choice on CD trials, as, during training, optimal choices were more similar to the probability only group than the frequency only group.

The likelihood ratings followed a similar pattern to choice preferences on AC trials in all three groups. If we assume that ratings reflect differences in expected value without influence of action reinforcement, it is difficult to attribute the choice effect on AC trials to an effect of action reinforcement alone. Considering the results on the whole, responding did not significantly differ between the probability x frequency group and the frequency-only group, indicating the increased probability associated with C in the probability x frequency group did not have a strong effect on preferences. Preferences in these groups therefore seem to be driven primarily by option frequency.

Experiment 2

Experiment 1 replicated the frequency effect on AC trials when A had a lower probability of the outcome than C, but had been presented more frequently than C. This occurred even with a greater difference in reward probability between A and C than that previously used (65 vs. 80 compared to 65 vs. 75 in Don et al., 2019). A novel finding in this experiment is that participants also rated option A as more likely to provide reward than option C. We assumed that these likelihood ratings would be less influenced by action reinforcement, as each option was presented individually, and participants were not required to make a choice. The similarity between choice and ratings may suggests that frequency effects are not strongly influenced by instrumental reinforcement, however we will return to this point in the General Discussion.

In Experiment 2, we aimed to determine whether the frequency effect persists when action reinforcement of alternative choice is removed from training. Estes (1976a, 1976b) has previously shown frequency effects in choice following observational training, which may indicate that action selection is not solely responsible for frequency effects. However, there is still reason to believe that action reinforcement might contribute to choice preferences in our task. In Estes' experiments, participants observe a hypothetical individual's preferred alternative between two stimuli in an opinion poll on each training trials, and then predict the likely preferred alternative in the same hypothetical population on transfer trials. In our task, participants receive point rewards for their choices, and so behaviour may be more strongly influenced by instrumental reinforcement. In addition, in observational studies, it is possible that participants could be covertly predicting a winner that is then reinforced when observing the outcome, in which case "A over B" is still reinforced to a greater extent than "C over D". Thus, we wanted to remove reinforcement of an option winning over an alternative, while participants are still actively participating in the task and receiving rewards for their choices. In Experiment 2, each card was presented individually during training, and participants were asked to pass or play each card. If participants chose to pass, the task would move on to the next trial without consequence (see Figure 5). If they chose to play, each option had the same probability of providing points as in Experiment 1. In this way, the task closely mimics Experiment 1, where the outcome of an option is not known if it is not chosen. By presenting the options individually, there will be no reinforcement of choosing A over the alternative during training, such that action reinforcement should not influence choice on AC trials at test. If action selection contributes to choice on AC trials, then we may expect a reduced frequency effect on AC trials. If choice is purely based on expected value, then we would still expect an effect of option frequency on choice preferences. The comparison groups were also included, with each option presented individually in their respective base-rates. If action selection is involved in frequency effects, we should expect a reduced effect wherever base-rates are manipulated, thus we should also see a reduced preference for A on AC trials in both the probability x frequency and frequency only group.

Method

Participants

The experiment received ethical approval from the Institutional Review Board (IRB) at Texas A&M University (IRB2019-0663D). One hundred and thirty-five undergraduates from Texas A&M University participated in return for partial course credit. In this experiment, we set a training criterion of playing the optimal cards (A and C) on average > 50% across training. All participants passed this criterion. Three participants had incomplete data for the reinforcement learning task, and were removed from the analyses, leaving 132 participants (89 female, mean age = 18.7, SD = 1.23). There were 44 participants in the probability x frequency group, 46 participants in the probability only group, and 42 in the frequency only group.

Procedure

The only difference between Experiment 1 and Experiment 2 was the presentation of options in the training phase. Here, participants were informed that they would be shown a card from one of four different decks on each trial, and would decide whether they would pass or play each card. They were told if they passed the card, the experiment would move on to the next trial. If they played the card, they had a chance of winning points. They were told some decks may be better than others for winning points. On each trial, one card appeared in the center of the screen, and the options "PASS" and "PLAY" were presented in rectangles beneath the card. Participants made their response by clicking on one of the options. If the participant chose to play, feedback was provided on the card, either "+10" for a reward trial, or 0 for a no reward trial, presented for 1500 ms. If the participant chose to pass the card, the card disappeared and the screen remained blank for 1500 ms, so that passing cards did not end the experiment sooner than playing cards. There was a 500 ms inter-trial interval for all trials. Participants were instructed that they could play as many cards as they liked. There were no penalties for either playing or passing cards. To ensure the task wasn't unnecessarily long, we included fewer B and D trials than A and C trials.

In Experiment 1, participants tended to choose the optimal cards more often, and therefore observed the outcome for B and D options to a lesser extent than A and C options. We therefore presented these trials less often than A and C trials, but retained a 2:1 base rate of both A:C trials, and B:D trials. In the probability x frequency and frequency only groups, there were 84 A trials, 56 B trials, 42 C trials, and 28 D trials. In the probability only group, there were 70 of each A and C trials, and 35 of each B and D trials. The remainder of the experiment continued in an identical manner to Experiment 1.

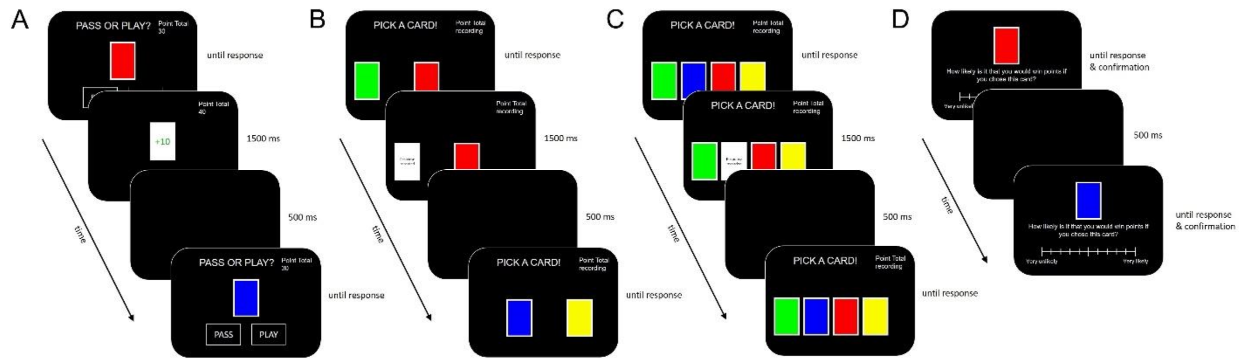

Figure 5. Schematic of the single cue version of the task.

Results & Discussion

Training

The proportion of play responses for each option in each group is shown in Figure 6. We again compared training performance between the probability x frequency and each of the control groups separately. As there were no penalties for passing or playing cards, and passing cards did not reduce the length of the experiment, the most optimal behavior would be to play the card on every trial. However, participants did show differences in the proportion of plays for different trial types, which is perhaps unsurprising as people often tend to respond in a suboptimal manner when outcomes are probabilistic, for example, probability matching (Neimark & Shulford, 1959; Newell & Shulze, 2016).

Probability x frequency vs. probability only. Participants were more likely to play the two more optimal options than the two less optimal options, $F(1,88) = 32.30$, $p < .001$, $\eta_p^2 = 268$, indicating learning of reward likelihoods, and this did not differ between groups, $F(1,88) = 0.10$, $p = .753$, $\eta_p^2 = .001$. Further analysis focused on the comparison of A and C trials. We ran a 2 x 2 x 7 mixed measures ANOVA with group as a between-subjects factor and trial type (A vs. C) and block (1-7) as within-subjects factors. This revealed a significant linear effect of block, indicating an increase in plays for the two optimal options across training, $F(1,88) = 16.84$, $p < .001$, $\eta_p^2 = .161$. There was also a significant main effect of trial type, indicating a greater proportion of plays for C than for A, $F(1,88) = 7.11$, $p = .009$, $\eta_p^2 = .075$, $BF_{10} = 12.51$, and this did not interact with group, $F(1,88) = 3.38$, $p = .069$, $\eta_p^2 = .037$, $BF_{incl} = 0.99$. There was also no main effect of group, $F(1,88) = 2.07$, $p = .153$, $\eta_p^2 = .023$, $BF_{10} = 0.49$.

Probability x frequency vs. frequency only. Participants were again more likely to play the two more optimal options than the two less optimal options, $F(1,84) = 34.44$, $p < .001$, $\eta_p^2 =$ 291, and this did not differ between groups, $F(1,84) = 3.52$, $p = .064$, $\eta_p^2 = .040$. Comparing A and C trials, there was again a significant linear effect of block $F(1,84) = 10.12$, $p = .002$, $\eta_p^2 =$ .108. However, there was no significant main effect of trial type, $F(1,84) = 1.47$, $p = .229$, $\eta_p^2 =$ .017, $BF_{10} = 0.25$, or group, $BF_{10} = 0.28$, $F(1,84) = 1.11$, $p = .294$, $\eta_p^2 = .013$. There appears to be a greater proportion of A than C plays early in training in the frequency only group than the probability x frequency group, and although there was no significant interaction between group and trial type, $F(1,84) = 3.36$, $p = .070$, $\eta_p^2 = .039$, the Bayes Factor was in favour of the alternative hypothesis, $BF_{incl} = 3.01$.

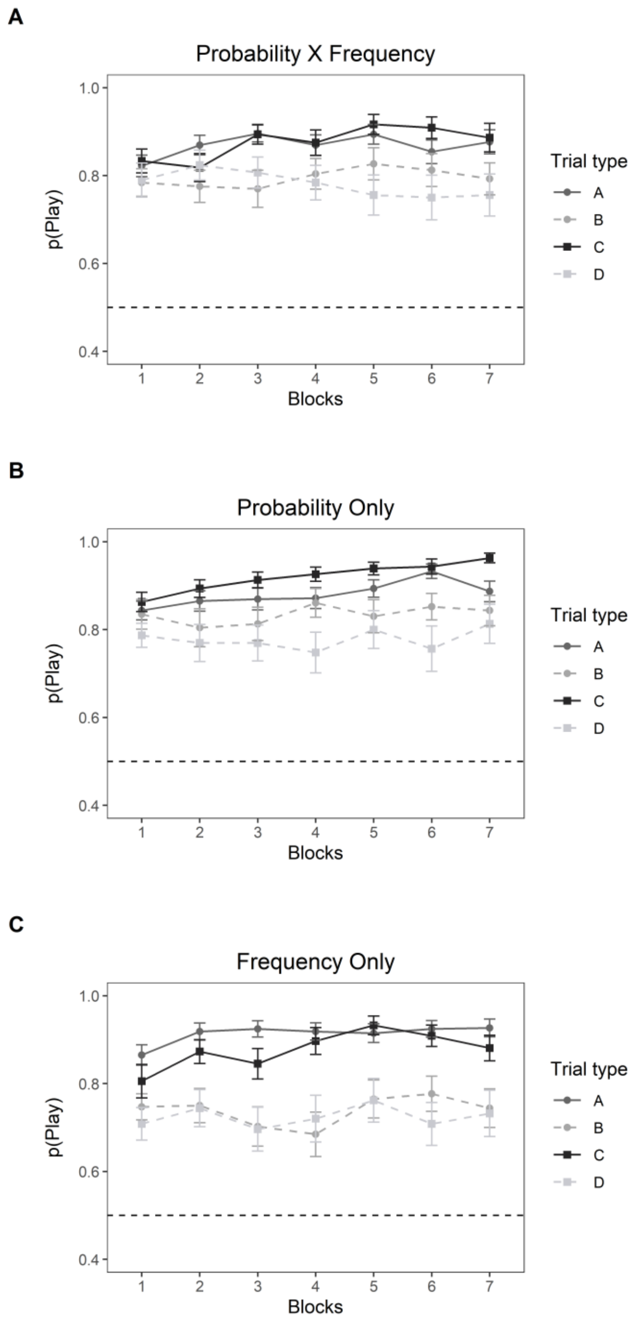

Figure 6. Proportion of plays for each option during training for a) probability x frequency group, b) probability group, c) frequency group in Experiment 2

AC Test Trials

The proportion of C choices on AC trials in each group is shown in Figure 7. In this case, the proportion of C choices did not significantly differ from chance in the probability x frequency group (*M* = .52, *SEM* = .05), *t*(43) = 0.35, *p* = .719, BF_{10} = 0.17. The proportion of C choices also did not differ from chance in the Frequency Only group (*M* = .46, *SEM* = .05), *t*(41) = -0.636, *p* = .528, BF_{10} = 0.20. Training cues individually therefore appears to remove the strong preference for option A seen in Experiment 1 when A is more frequent. In the probability only group, participants responded optimally, with a significant preference for option C (*M* = .66, *SEM* = .06), *t*(45) =3.61, *p* = .001, BF_{10} = 37.35. Although there was no significant preference for A in the probability x frequency group, there was still an effect of option frequency, as there were significantly fewer optimal choices in the probability x frequency group compared to the probability only group, *t*(88) = 2.23, *p* = .028, BF_{10} = 1.90. There was no significant difference in responding between the probability x frequency and frequency only groups, *t*(84) = 0.73, *p* = .470, BF_{10} = 0.28. Responses for all other test trials are shown in supplementary material.

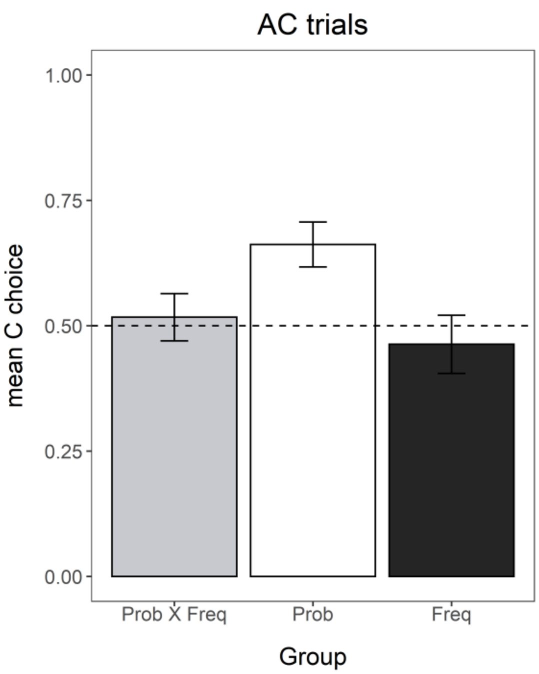

Figure 7. Proportion of C choices on AC trials in the transfer test in Experiment 2. The dashed line indicates chance.

Ratings

Mean likelihood ratings are shown in Figure 8. Comparing probability x frequency and probability only groups, there was a significant main effect of trial type, *F*(1,88) = 6.94, *p* = .010, η_p^2 = .073, BF_{10} = 11.84, where ratings were higher overall for C than for A, and no interaction with group, *F*(1,88) = 1.32, *p* = .254, η_p^2 = .015, BF_{incl} = 0.44. Comparing probability x frequency and frequency only groups, there was no effect of trial type, *F*(1,84) = 0.04, *p* = .845, η_p^2 < .001, BF_{10} = 0.17. While there appears to be slightly higher ratings for C than A in the probability x frequency group, and slightly higher ratings for A than C in the frequency only group, there was no significant interaction between trial type and group, *F*(1,84) = 1.34, *p* = .251, η_p^2 = .016, BF_{incl} = 0.49. In both training and test results in Experiment 2, we found potentially meaningful, but non-significant trends that may reach significance with a larger, higher-powered sample size. However, in this case, the Bayes Factor indicated more support for the null hypotheses that there was no interaction.

Overall in Experiment 2, both frequency groups were affected by removing choice between alternatives in training, as there was no longer a strong preference for A. However, there was still an effect of option frequency, as responses differed between the probability x frequency and probability only groups. This suggests that action reinforcement may be a contributing factor to the strength of frequency effects, but is not the sole cause of frequency effects in reinforcement learning. Ratings were again largely consistent with the pattern of choice preferences, which is discussed in the General Discussion.

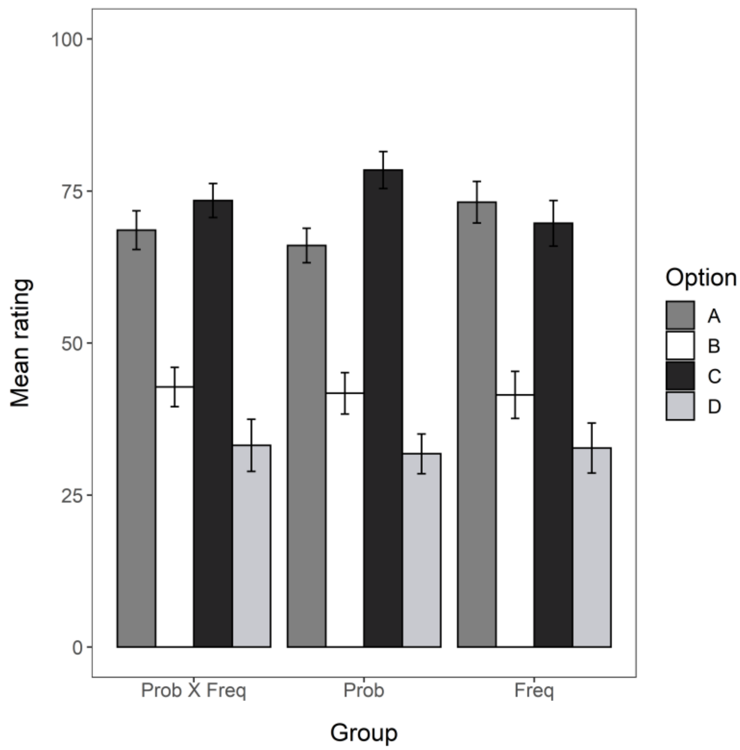

Figure 8. Mean ratings of the likelihood of receiving reward for each option in Experiment 2.

Computational modeling

While we observed effects of option frequency in both Experiment 1 and Experiment 2, there is an apparent difference in the strength of the frequency-effect depending on training conditions, which might indicate a role of reinforcing action selection. Delta and Decay reinforcement models update the expected values of options, and the probability of choice of each option is based on these expected values. They do not necessarily reflect an instrumental process where the choice itself is reinforced. Before making strong conclusions about the role of action selection in these effects, it is important to understand whether these learning models can account for the difference between Experiments, and in particular, whether the Decay model is still the best performing model across these two procedures. The models should operate similarly regardless of whether options are presented in binary pairs, or alone. That is, the Delta model should still give greater value to the higher probability option, and the Decay model should give greater value to the more frequent option in both training procedures. We therefore fit the data from each task with Delta and Decay models. Here we focused on the basic models, which each contain two free parameters. Including additional parameters may allow for better model fit, yet these more parsimonious models make diverging predictions when there are differences in option frequency, which allows for strong inference as there are possible patterns of human behaviour that each model cannot account for (Platt, 1964; Roberts & Pashler, 2000). See Don et al. (2019) for an evaluation of several extended models in the choice version of this task.

For the Delta model, expected values (EVs) for each *j* option were calculated as:

$$EV_j(t+1) = EV_j(t) + \alpha \cdot (r(t) - EV_j(t)) \cdot I_j \tag{1}$$

Where r is 1 if a reward is received on that trial, and 0 otherwise. Ij is an indicator variable coded as 1 if option *j* was chosen, or played, on that trial, and 0 otherwise. This means that expected values are only updated for the option chosen on each trial. The prediction error (r(t) – EV(t)) specifies the difference between what occurs on trial t, and what is expected, and is modulated by a learning rate parameter ($0 \leq \alpha \leq 1$). Higher values of α indicate greater weight to recent outcomes, while lower values indicate less weight to recent outcomes.

The Decay rule updates expected values for each j option according to:

$$EV_j(t+1) = EV_j(t) \cdot (1 - A) + r(t) \cdot I_j \tag{2}$$

where *A* is a decay parameter. We used *(1-A)* as the decay parameter so that higher values of *A* indicate more decay and lower values indicate less decay, such that they are more comparable to the learning rate parameter in the Delta model. In both models, higher values indicate a greater reliance on recent outcomes. As in Equation 1, I_j is an indicator variable that is set to 1 if option *j* was selected on trial *t*, and 0 otherwise. This means that EVs will increment by the reward for the chosen option only, but all options will decay on every trial.

At test and in the choice version of the task, the predicted probability that option *j* will be chosen on trial *t*, $P\left|C_j(t)\right|$ was determined by entering EVs into a Softmax rule:

$$P\left|C_j(t)\right| = \frac{e^{\beta \cdot EV_j(t)}}{\sum_1^{N(j)} e^{\beta \cdot EV_j(t)}} \tag{3}$$

Where $\beta = 3^c - 1$ ($0 \leq c \leq 5$), and *c* is a log inverse temperature parameter that determines how consistently the option with the higher expected value is selected (Platt, 1964; Roberts & Pashler, 2000). Lower values of c indicate more random choices, and higher values indicate more deterministic choices, where the option with the highest expected value is selected most often. Defining β in this way allows it to take on a very large range of values (0-242), and is equivalent to setting a prior on beta with a truncated exponential distribution.

For the single cue version of the task, we used a cumulative distribution function to calculate the probability of playing the card on each trial during:

$$(P(play) = P(X < \beta \cdot (EV_j - B)) \tag{4}$$

Where *B* is a threshold parameter that determines the limit for deciding to play a card, and β is a scaling parameter similar to the inverse temperature parameter, where $\beta = 3^t - 1$ ($0 \leq t \leq 5$). This parameter varied independently from the inverse temperature parameter in the Softmax rule. The expected value of each option was then updated only if the card was played. The probability of choosing each option at test was determined using the Softmax function in Equation 3.

Simulations

To verify general model predictions for each group, we simulated each model on the choice and single cue versions of the task. We ran 100 simulations of each parameter combination, each with randomised trial order. In order to best demonstrate differences in choice proportions based on expected value, we held the inverse temperature parameter for the Softmax function (c) constant at 2.5, but averaged over all other values of the remaining parameters.

Table 2 summarises the predicted probability of choosing C on AC trials for each group, averaged across those parameter combinations, as well as mean expected values for each option produced by each model by the end of the training phase. The Delta model bases value on the probability of reward provided by each option, and predicted preferences for C were consistent with this in both choice and single cue versions of the task. The model predicted a bias towards C in the probability only and probability x frequency conditions, and no bias in the frequency only condition, where there were no differences in the probability of reward provided by A and C. These predictions were fairly consistent across both versions of the task. The strength of the predicted bias to C was weaker in the single cue version of the task, but there was still a clear preference for the higher probability option.

In comparison, the Decay model bases value on the cumulative rewards provided by each option, and therefore more frequent options are given higher value. The model predicted a preference for the more frequently rewarded option in all three groups. This is option A in probability x frequency and frequency only groups, and C in the probability only group. When base-rates are equal, the higher probability option will provide a greater number of rewards. The Decay model's predictions were also generally more extreme in the choice version of the task than the single cue version of the task. Thus, both models anticipate some difference in the degree of effects across the different training conditions, but still predict preferences in opposing directions when frequency is manipulated. On the whole, human behaviour is more consistent with the Decay model, particularly for the choice version of the task.

Table 2.

Mean simulated C choices on AC trials, and simulated expected values for each choice option

Model	Experiment	Group	Mean C on AC	EV A	EV B	EV C	EV D
Delta	Choice	Prob X Freq	0.66	0.56	0.18	0.73	0.19
		Prob only	0.67	0.56	0.19	0.74	0.18
		Freq only	0.50	0.62	0.17	0.62	0.20
	Single cue	Prob X Freq	0.59	0.57	0.40	0.66	0.32
		Prob only	0.60	0.57	0.40	0.66	0.32
		Freq only	0.50	0.60	0.37	0.60	0.38
Decay	Choice	Prob X Freq	0.38	5.04	1.18	3.63	0.22
		Prob only	0.60	3.42	1.10	4.94	0.43
		Freq only	0.32	5.54	0.95	2.76	0.50
	Single cue	Prob X Freq	0.41	3.31	1.19	2.04	0.36
		Prob only	0.55	2.75	0.76	3.39	0.45
		Freq only	0.39	3.57	1.03	1.78	0.53

Model fits.

We also fit each model to participants training and test data, and to the test data alone after training the models on training trials, using the optim function in R. Model fits were compared using the Bayesian Information Criterion (BIC; Schwarz, 1978). BIC differences (ΔBIC = Delta BIC – Decay BIC) were also transformed into a Bayes Factor representing the evidence that the Decay model is the better model, calculated as exp $\left(\frac{\Delta \text{BIC}}{2}\right)$ (Wagenmakers, 2007).

Table 3.

Model comparisons including average BIC, ΔBIC, Bayes Factors and Pseudo R^2s

			BIC				Pseudo R^2	
Fit to	Experiment	Group	Delta BIC	Decay BIC	ΔBIC	BF	Delta	Decay
All trials	Choice	Prob X Freq	329.72	325.00	4.72	10.58	0.15	0.16
		Prob only	284.99	280.88	4.11	7.82	0.27	0.28
		Freq only	322.88	319.89	2.99	4.47	0.16	0.17
		total	312.56	308.65	3.92	7.09	0.19	0.20
	Single cue	Prob X Freq	314.82	315.24	-0.42	0.81	0.36	0.36
		Prob only	286.34	280.96	5.38	14.70	0.42	0.44
		Freq only	297.02	293.03	3.99	7.33	0.40	0.41
		total	299.23	296.23	3.00	4.48	0.40	0.40
Test trials	Choice	Prob X Freq	147.64	143.97	3.67	6.26	0.17	0.19
		Prob only	135.00	130.89	4.11	7.82	0.25	0.27
		Freq only	153.75	149.76	3.99	7.34	0.13	0.16
		total	145.59	141.67	3.93	7.12	0.18	0.21
	Single cue	Prob X Freq	157.47	157.95	-0.48	0.79	0.17	0.17
		Prob only	154.23	149.16	5.07	12.65	0.19	0.22
		Freq only	151.05	148.00	3.06	4.61	0.21	0.23
		total	154.30	151.72	2.58	3.63	0.19	0.20

Table 4.

Model comparisons including average BIC, ΔBIC and Bayes Factors by condition

Fit to	Group	Delta BIC	Decay BIC	ΔBIC	BF
All trials	Prob X Freq	321.42	319.57	1.85	2.53
	Prob only	285.75	280.93	4.82	11.14
	Freq only	309.30	305.79	3.51	5.80
	total	305.26	301.845	3.41	5.52
Test trials	Prob X Freq	153.11	151.76	1.36	1.97
	Prob only	145.79	141.14	4.65	10.24
	Freq only	152.33	148.84	3.50	5.75
	total	150.36	147.17	3.19	4.93

Table 3 shows the average BIC, ΔBIC, and Bayes factors for each model for each group, and for all participants. When fit to all trials in Experiment 1, the Decay model provided a better fit of the data overall (ΔBIC = 3.92, BF = 7.09), and for each group individually. Interestingly, in the probability only group where relative estimated values based on reward probability and cumulative reward should be similar, the Decay model provided a better fit to the data than the Delta model. In Experiment 2, both models provided a similar fit to the critical probability x frequency group, suggesting that neither model better accommodated the choice effect in this condition. This could simply be due to the fact the models each predict a clear preference in choice for one option or the other, while participants showed no preference in choice, and therefore the human data fall somewhere in the middle of the two model predictions. The pattern of fits was largely the same when the models were fit to the test trials only (see Table 3). We also compared each model to a random or null model by computing McFadden's pseudo R^2 (McFadden, 1973). When fit to all trials, both models show an improvement in fit over a null model, with pseudo R^2 = .19 and .20 for Delta and Decay models, respectively in Experiment 1, and pseudo R^2 = .40 for both models in Experiment 2. When fit to the test trials, R^2 = .18 and .21 for Delta and Decay models, in Experiment 1, and R^2 = .19 and .20 for Delta and Decay models in Experiment 2. The large R^2 for both models in Experiment 2 fit to all trials compared to when fit to test trials only might suggest the models are able to predict training performance particularly well. We also compared model fits by condition, collapsed across experiments. Mean BICs and BFs for each condition are shown in Table 4. The Decay model provided a better fit for each of the conditions, collapsed across experiments.

Ex-post simulations

Additional simulations were run to determine how well each model could reproduce the AC choice effects shown in the human choice data, using the best-fitting parameters (Palminteri et al., 2017). The simulations were run twice, once using participants' best-fitting parameters fit to the entire experiment, (post-hoc simulations), and once fit to the training phase only (a priori simulations; Ahn et al., 2008; Busemeyer & Wang, 2000). These best-fitting parameters were used to generate predictions for the entire data set (see Supplementary Material for average best fitting parameters for each condition in each experiment). For Experiment 2, the *c* parameter in the Softmax rule is only used at test, thus when simulating data using the best-fitting parameters fit to the training phase only, we used the average *c* parameter for each group when fit to the entire experiment. For each experiment, we ran 1000 simulations for each group, sampling with replacement from the relevant participants' best fitting parameters from each model. Average predicted proportion of C choices are shown in Figure 9 for simulations based on best fitting parameters from the entire experiment, and Figure 10 for simulations based on best fitting parameters from training only. Neither model fully predicted the pattern of results across Experiment 1 and Experiment 2. The Delta model predicted a preference for C in the probability x frequency group in both experiments, and no effect of frequency in Experiment 2, while the Decay model continued to predict a frequency effect in both Experiments.

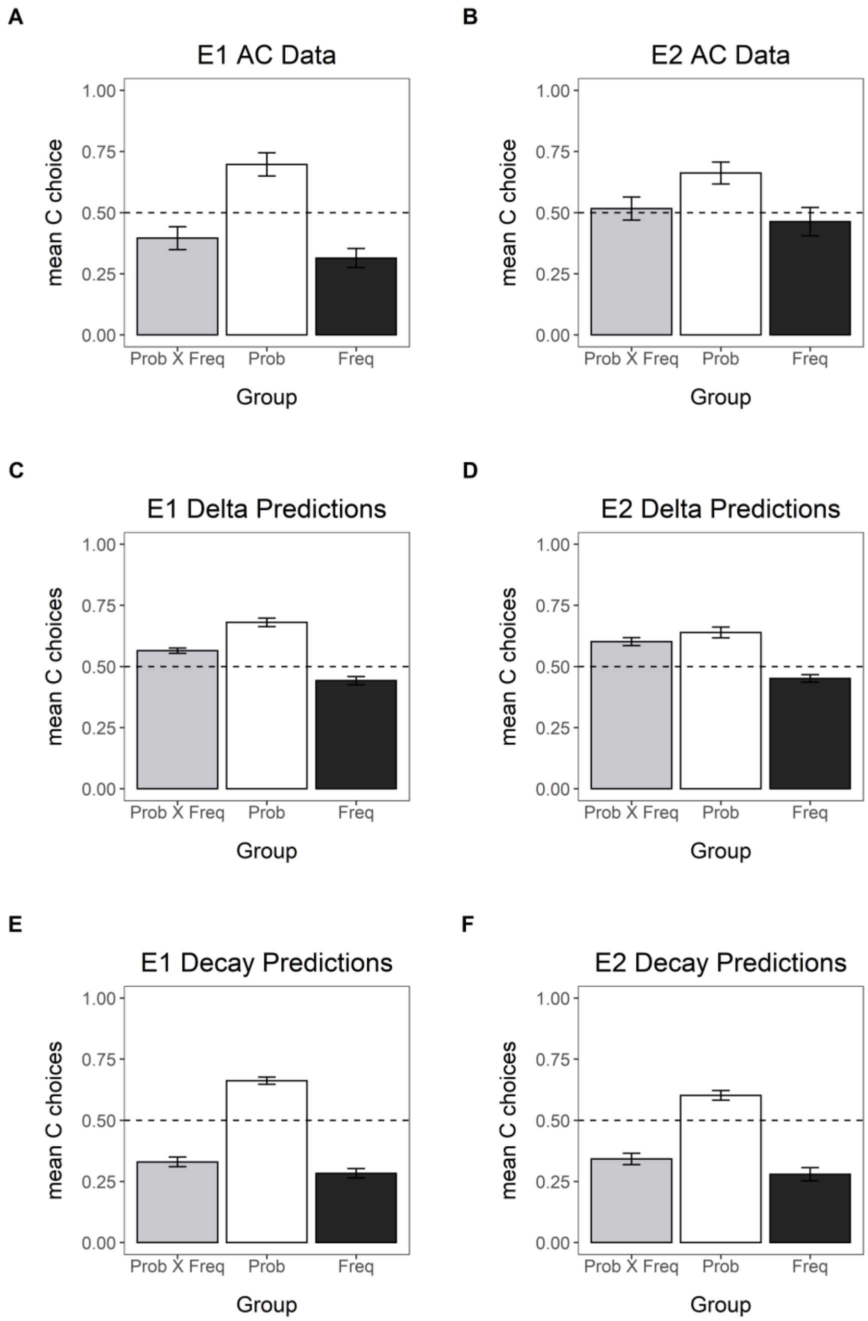

Figure 9. Observed and simulated predictions for AC trials at test, using the best fitting parameters from each group, fit across the entire experiment.

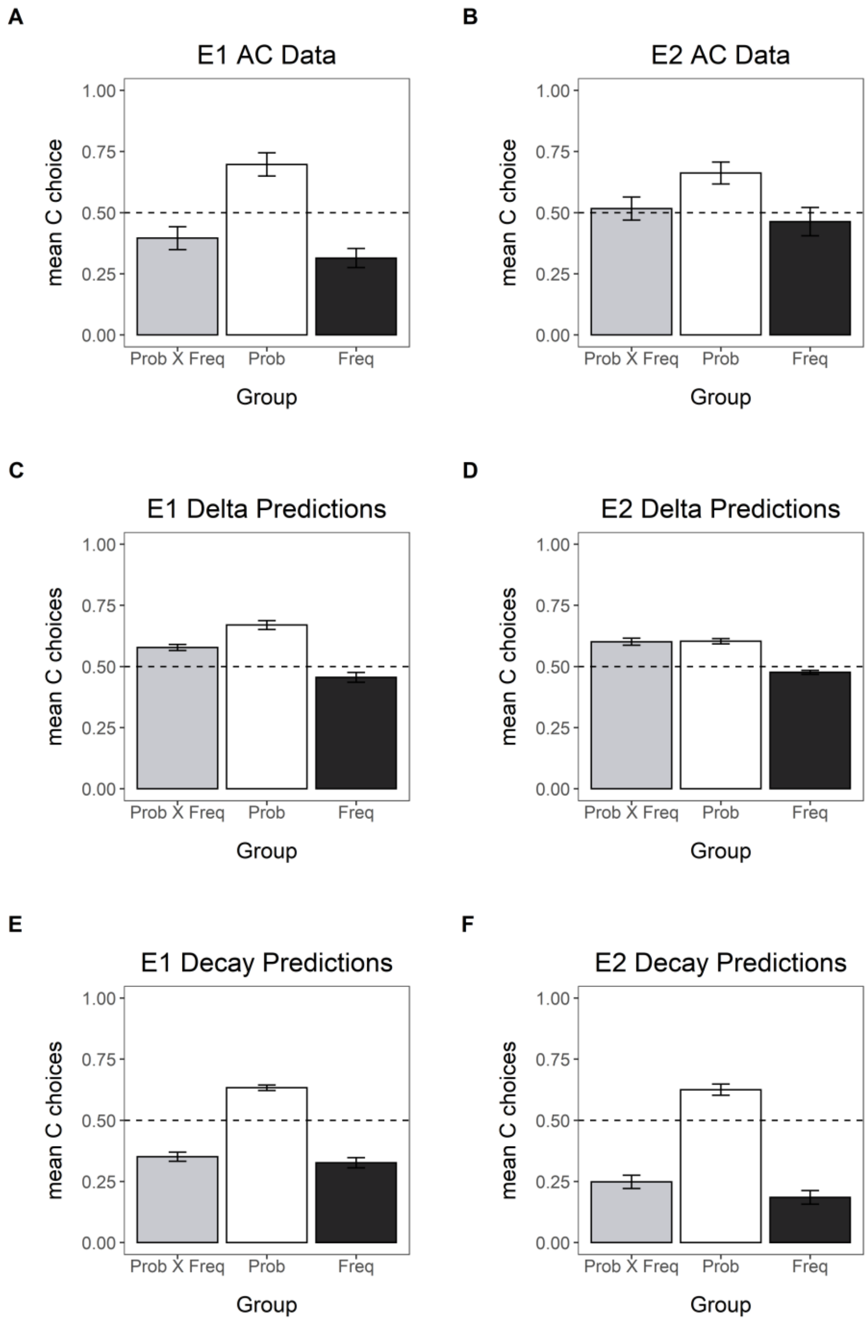

Figure 10. Simulated predictions for AC trials at test, using the best fitting parameters from each group, fit to training trials only.

General Discussion

This study aimed to assess the effect of option frequency in reinforcement learning. In particular, we aimed to test whether action selection plays a contributing role in the preference for frequently experienced choice options in a reinforcement learning task. In two experiments, we demonstrated an influence of option frequency on both choice preferences and expectations of reward likelihood. In both experiments, in the probability x frequency group, option A had a lower probability of reward than option C, but was experienced more frequently. In Experiment 1, where the payoff provided by options A and C were learned through choice responses in separate pairs, participants favored the higher frequency option A over C, even though it had a lower probability of reward, replicating the findings of Don et al. (2019). They also preferred the more frequent option in the frequency only group, where A and C had the same probability of reward, and A was more frequent than C, replicating the findings of Estes (1976b). In Experiment 2, the preference for A was effectively removed in both probability x frequency and frequency only groups when cues were presented individually during training. The absence of an A-preference in these groups is perhaps the strongest evidence for a role of action reinforcement to the effect. However, there was still a clear influence of option frequency on choice preferences, as there was a significant attenuation of the preference for C compared the probability only condition, which showed consistent preferences for the high probability option C. Thus, while instrumental reinforcement might contribute to the strength of the effect, it cannot completely account for the effect of option frequency. However, a full interpretation of the results requires some consideration of the likelihood ratings phase.

A critical novel finding of this study is that participants rated option A as more likely to provide reward than option C in Experiment 1 when A was experienced more frequently than C, in both the probability x frequency and frequency only groups. Thus, option frequency appears to affect beliefs about the likelihood of receiving reward. This may have important implications for understanding people’s persistence in pursuing frequently experienced sources of reward, even if the likelihood of reward is low (e.g., cheap but largely ineffective herbal remedies, or problem gamblers playing slot machines). Such results may be a reflection of strong, automatically coded memory for frequency (see Ekstrand et al., 1966; Hintzman, 1988), which may then influence likelihood judgments. We initially assumed that these ratings would be unlikely to be directly affected by action reinforcement in this task. Unlike the AC choice trials, each option was presented individually, and no alternate choice judgment was made. Overall, the data suggest there is little dissociation between tests that we assume do and do not require action selection. Instead, choice preferences and ratings were largely consistent; in Experiment 1, where we observed a choice effect in ratings towards A on AC trials when A was more frequent (in both the probability x frequency and frequency only groups), we also saw a bias where ratings were higher for A than for C. In Experiment 2, where there was no preference for A in choice in either the probability x frequency or frequency only group, there was also no preference in ratings.

Considering both experiments, the results are consistent with the idea that instrumental conditioning is playing some role in the bias towards A. When action reinforcement is removed from training, frequency appears to affect the expected value of choice alternatives, reducing preferences for the higher probability, less frequent option C. The addition of action reinforcement in choice training in Experiment 1 may then further bias choice towards A over C. Here, the frequency of AB trials would strengthen conditioning of the action of choosing A to a greater extent than that of choosing C, such that participants are more likely to choose A on AC trials. How then do we reconcile this explanation with the results from the ratings test phase? If the differences in choice tests between the two experiments are attributable to action selection, and if ratings are not influenced by action selection processes, then we would expect the rating phase to be unaffected by differences in training conditions, and remain similar across both experiments. Instead, the pattern of ratings was consistent with the pattern of choice within each experiment. We therefore need to consider what ratings are reflecting. There are several interpretations we can make here. The first is that ratings are reflecting something other than expected value. One possibility is that participants are simply attempting to respond consistently across test phases. For instance, if participants chose A on AC trials, they may justify this response by rating the likelihood of reward for this option as high. Similarly, if they show less preferences in choice, they may rate the options as having similar likelihood of reward. Future research could control for this by counterbalancing the order of test phases, or separating the tests between-subjects.

The second interpretation is that action reinforcement does not play a role in the frequency effect, and both choice preferences and ratings are a reflection of expected reward alone. However, if this were the case, expected reward would be similar between Experiment 1 and Experiment 2, so this explanation is difficult to reconcile with the reduced preference for A in the frequency groups in Experiment 2, and is also not well supported by the fits and simulations of reinforcement learning models that estimate expected value but do not appeal to action selection. Initial model simulations did predict reduced effects in the single cue version of the task than the choice version, but both models still predicted a clear preference in opposing directions. While the Decay model provided a better fit to the data overall, neither model provided a better fit to the probability x frequency group in Experiment 2, and neither model adequately reproduced the differences in the frequency-effect on AC trials between Experiment 1 and Experiment 2 using the best-fitting model parameters. This suggests that expected value – at least those assumed by these models – is insufficient to explain the results on the whole.

The third, and our preferred interpretation is that action reinforcement and expected value are not separable in the way we have assumed here, and reinforcing choice during training may also influence expected value in a way not captured by the Delta and Decay models. That is, expected values update based on reward, and in addition, the act of choosing an option over an alternative further increases its perceived value in such a way that not only increases the probability of it being chosen in the future, but also increases the perception that it is likely to provide reward. In this case, we might expect more extreme differences in value following choice training in Experiment 1 than single cue training in Experiment 2, which would lead to both greater choice and ratings of A over C in Experiment 1 than Experiment 2 as well as similar patterns of responding across choice and ratings in both experiments. This interpretation is therefore most consistent with the results we observed.

Our results contrast somewhat with those of Estes (1976a, 1976b), who found preferences for the more frequent option following observational training, which should not involve instrumental reinforcement of choice. It could simply be the case that instrumental reinforcement plays a more significant role when people actively make choices that are rewarded, such that it increases the strength of the effect. On the other hand, perhaps the critical factor is the act of comparing options during training. That is, in Estes' tasks, observing A win over B more frequently than C win over D may covertly reinforce "A wins", in a similar way to direct reinforcement of choosing A over B. This may drive a stronger preference to A on AC trials than when its associated outcome is observed separately during training. As mentioned above however, this is not the sole cause of the effect, as we still see an influence of option frequency in the comparison between the probability x frequency and probability only groups.

An alternative explanation for the difference between Experiment 1 and Experiment 2 is that training options in separate AB and CD pairs may hinder comparison between A and C, such that it may be more difficult for participants to assess the difference in probability of reward they provide, or they are discouraged from making this comparison as the within-trial comparison may be more salient. Training cues individually may then allow greater comparison between A and C trials, leading to a greater number of optimal responses on AC trials. However, if it were the case that separate training hinders learning that C is more optimal, then we should also expect poorer performance in the probability-only group in Experiment 1 where A and C were also experienced on separate choice trials. Yet, this group showed more optimal choices on AC trials, and did so to a similar magnitude in both experiments.

A different interpretation of frequency effects in this task is that uncertainty about option C drives choice towards the more frequently experienced – and therefore more certain – option A. We should therefore consider whether different training conditions influence uncertainty about option outcomes. It is possible that the single cue design reduces uncertainty about the option outcomes, as participants are given the opportunity to see the outcome for an option on every single trial, if they choose to play each card. In comparison, in choice training, if we assume some level of exploration of choices during learning, there will be some uncertainty about the potential outcome of the unchosen option on each trial. This is then perhaps consistent with the reduced preference for A in Experiment 2, as individual exposure to C might reduce uncertainty about that option. However, this explanation doesn't necessarily speak to the *relative* uncertainty of A compared to C, which, given the matched base-rates, should be consistent across experiments. That is, between experiments, participants have the same opportunities to learn about the outcome associated with A and C, which might instead suggest no differences in relative uncertainty between conditions. Further understanding the role of uncertainty in this task is an important focus for future research, and will require designs that can adequately manipulate uncertainty.

Overall, the current study demonstrates a clear effect of option frequency on decision making and some contribution of action reinforcement to this effect. Similar effects were found decades ago by Estes (1976a; 1976b), and these results have implications for models that have been developed over the past several decades, such as delta models, which have difficulty accounting for frequency effects in reinforcement learning tasks. We found that presenting cues individually during training reduced preferences for the more frequent option, and the difference between training conditions was not well anticipated by reinforcement learning models. However, the effect of option frequency cannot be completely accounted for by action reinforcement. While removing choice from training reduced the strength of the preference for A, we still observed an effect of option frequency on choice, as there were fewer optimal choices than a comparison group with equal base-rates. We also found little dissociation between a choice test phase that should be influenced by action selection, and a ratings test that we assumed would be unaffected by this process. The results instead suggest that instrumental reinforcement influences both choice of an option and judgements about its rewarding properties. The effect of frequency on judgments of the likelihood of receiving reward is novel, and indicates that the frequency with which we experience reward related stimuli influences not only choice, but also beliefs that may maintain suboptimal decision making. On the whole, the results suggest that action reinforcement may influence the value of each option during training. It remains to be seen whether there are similar effects of option frequency when losses are involved, with continuous rewards, or if these effects extend beyond reward scenarios into other types of learning (e.g., category learning, causal learning etc.), which are important avenues for future research.

Acknowledgements

This work was supported by NIA grant AG043425 to DAW. The authors declare no conflicts of interest. Data and model code are available on the Open Science Framework (OSF): https://osf.io/7yzgf/

References

Ahn, W., Busemeyer, J.R., Wagenmakers, E., & Stout, J.C. (2008). Comparison of decision learning models using the generalization criterion method. *Cognitive Science, 32*, 1376-1402. doi: 10.1080/03640210802352992

Barto, A.G. (1992). *Reinforcement learning and adaptive critic methods*. In Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches. D.A. White & D.A. Sofge, Eds.: 469–491. Van Norstrand Reinhold: New York.

Barto, A.G. (1995). *Adaptive critics and the basal ganglia*. In Models of Information Processing in the Basal Ganglia. J.C. Houk, J.L. Davis & B.G. Beiser, Eds.: 215–232. MIT Press: Cambridge, MA.

Brainerd, C. J. (1981). Working memory and the developmental analysis of probability judgment. *Psychological Review, 88*, 463. doi: 10.1037/0033-295X.88.6.463

Busemeyer, J.R., & Wang, Y.M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44,* 171-189. doi: 10.1006/jmps.1999.1282

Don, H., & Worthy, D. A. (2020, May 5 28). Frequency effects in action versus value learning. Retrieved from osf.io/7yzgf/

Don, H. J., Otto, A. R., Cornwall, A. C., Davis, T., & Worthy, D. A. (2019). Learning reward frequency over reward probability: A tale of two learning rules. *Cognition, 193*, 104042. doi: 10.1016/j.cognition.2019.104042

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgement and choice. *Annual Review of Psychology, 32,* 53-88. doi: 10.1146/annurev.ps.32.020181.000413

Erev, I., & Roth, A.E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review, 88*, 848-881. doi: 10.1006/game.1999.0783

Estes, W.K. (1976a). The cognitive side of probability learning. *Psychological Review, 83,* 37-64. doi: 10.1037/0033-295X.83.1.37

Estes, W. K. (1976b). *Some functions of memory in probability learning and choice behavior*. In Psychology of Learning and Motivation (Vol. 10, pp. 1-45). Academic Press. doi: 10.1016/S0079-7421(08)60463-6

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?. ECVP Abstract Supplement, *Perception, 36*.

McFadden, D. (1973). *Conditional logit analysis of qualitative choice behavior*. In P. Zarembka (ed.), Frontiers in Econometrics. New York: Academic Press: 105-142.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy, 28,* 487-495. doi: 10.1016/0005-7967(90)90135-6

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences, 1104*, 35-53. doi: 10.1196/annals.1390.022

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences, 21*, 425-433. doi: 10.1016/j.tics.2017.03.011

Patrick, C. J. (2010). *Operationalizing the triarchic conceptualization of psychopathy: Preliminary description of brief scales for assessment of boldness, meanness, and disinhibition.* Unpublished manuscript. University of Minnesota. Minneapolis, MN

Patrick, C. J., Kramer, M. D., Krueger, R. F., & Markon, K. E. (2013). Optimizing efficiency of psychopathology assessment through quantitative modeling: Development of a brief form of the Externalizing Spectrum Inventory. *Psychological Assessment, 25,* 1332. doi: 10.1037/a0034864

Platt, J.R. (1964). Strong inference. *Science, 146,* 347-353. doi: 10.1126/science.146.3642.347

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385-401. doi: 10.1177/014662167700100306

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A.H. Black & W.F. Prokasy (Eds.) *Classical conditioning II: Current research and theory.* New York: Appleton-Century-Crofts.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358-367. doi: 10.1037/0033-295x.107.2.358

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E. J. (2017). Bayesian analysis of factorial designs. *Psychological Methods, 22,* 304. doi: 10.1037/met0000057

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461-464. doi:10.1214/aos/1176344136

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). Manual for the state-trait anxiety scale. *Consulting Psychologists Press*.

Sutton, R.S., & Barto, A.G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT.

Thorndike, E. L. (1911). *Provisional laws of acquired behavior or learning.* Animal Intelligence. The Mc Millian Company: New York.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language*, *28*, 127-154. doi: 10.1016/0749-596X(89)90040-5

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063. doi: 10.1037//0022-3514.54.6.1063

Wagenmakers, E.J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14,* 779-804. doi: 10.3758/BF03194105

Wagenmakers, E.J. et al. (2018). Bayesian inference for psychology Part II: Example applications with JASP. *Psychonomic Bulletin and Review, 25*, 58-76. doi: 10.3758/s13423-017-1323-7

Widrow, B., & Hoff, M.E. (1960). Adaptive switching circuits. *1960 WESCON Convention Record Part IV,* 96-104.

Williams, R.J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning, 8,* 229-256. doi: 10.1007/BF00992696

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., Hink, R. F., & Reed, B. (2011). Improving digit span assessment of short-term verbal memory.

Journal of Clinical and Experimental Neuropsychology, *33*, 101-111. doi: 10.1080/13803395.2010.493149

Yechiam, E., & Busemeyer, J.R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision-making. *Psychonomic Bulletin & Review, 12,* 387-402. doi: 10.3758/BF03193783

Yechiam, E. & Ert, E. (2007). Evaluating the reliance on past choices in adaptive learning models. *Journal of Mathematical Psychology*, *51*, 75-84. doi: 10.1016/j.jmp.2006.11.002