Frequency Effects in Decision-Making are Predicted by Dirichlet Probability Distribution Models

Astin Cornwall, Hilary Don, & Darrell Worthy
Texas A&M University, Texas, USA

Abstract

Frequency of reward and average reward value are two types of reward information we utilize when making decisions between two alternative options. Often, these two pieces of information coincide with the highest value option, however, when a slightly less valuable option is presented more frequently, standard reinforcement learning models such as the Delta model can make incorrect predictions. This paper explores the discrepancy in these predictions by way of simulating relevant behavioral tasks with the Delta model, the Decay model, and a novel Bayesian model based on the Dirichlet distribution. We then compare model predictions to behavioral data from some of the same tasks that were simulated. The Delta model provides a poor fit to the data for each of the three presented tasks when compared to the Decay model and the two Bayesian learning models, because it predicts a bias toward options with higher average reward, while the Decay and Bayesian models predict a bias toward reward frequency. The Decay and Bayesian models show a distinct similarity in prediction and fits to the data for most of the tasks. This is because both models predict a bias toward reward frequency rather than average reward magnitude, despite different computational formalisms. However, we also note some interesting discrepancies between the Decay and Bayesian models which will show that in some cases, the frequency of reward may be more important than the reward value.

Keywords: Frequency Effect; Reinforcement Learning; Bayesian Learning

Introduction

A wide variety of decisions we make on a day-to-day basis are repetitive in the sense that we may choose one option over another fairly consistently. Whether these decisions are about choosing name-brand over store-brand items, restaurant A or B, or taking the freeway versus the side roads to work, it's possible that all of these decisions are computed by common algorithmic mechanisms. These decisions could be based on the average outcome of each option, for example, taking the freeway to work is nearly always faster than taking the side roads. However, supposing that a new bypass opens that is predicted to greatly reduce travel time, a person may still be inclined to choose the freeway since they have had many more experiences with the freeway being adequate enough.

Learning rules in formal models of cognition allow us to make sense of human decision-making processes and get a glimpse as to why people make the decisions they do in situations such as the examples above. In this paper, we compare the choice predictions of four learning models: the Delta rule, the Decay rule, and two Dirichlet distribution-based models, on a set of decision-making tasks.

The Delta rule, in particular, is a widely used learning rule across many domains of cognition. This model predicts that people will have a preference for options that have the greatest expected value, based on representations of the *average* reward for each option, amongst alternative options (e.g. Busemeyer & Stout, 2002; Daw et al., 2006; Gluck & Bower, 1988; Jacobs, 1988; Rescorla & Wagner, 1972; Rumelhart & McClelland, 1986; Sutton & Barto, 1981; 1998; Widrow & Hoff, 1960; Williams, 1992).

In contrast to the Delta rule's average value representations, the Decay rule learns to represent the *cumulative* value of each option based on the frequency with which it has been rewarded. Psychologically, the Decay model assumes that that decision outcomes are stored in working memory and decay over time. The Decay model utilizes a decay parameter which diminishes the expected value of each option at each timepoint. Therefore, the option with the greatest expected value in this model would be the option which is most frequently rewarded; in most cases (Erev & Roth, 1998; Yechiam & Busemeyer, 2005; Yechiam & Ert, 2007).

In a departure from these two standard learning models, this paper presents a Bayesian model which simply learns how many times each option has a positive outcome rather than learning expected values. The Dirichlet Probability Distribution (DPD) model holds in memory a representation of how many times each option has produced a reward, regardless of value. Each of these values are used as the concentration parameter values in the distribution which allocates more probability mass to the options which have been rewarded most frequently. Thus, when attempting to choose between options, the option more frequently rewarded will have a higher probability of being chosen.

The sole use of the Dirichlet distribution as the base for this model may seem atypical considering it is more often used in Bayesian data analysis for determining the clustering, or categorization, of data (e.g. Griffiths, Sanborn, Canini, & Navarro, 2008), or in Dirichlet Process or Mixture models (Navarro, Griffiths, Steyvers & Lee, 2006; Gershman & Blei, 2012; Sims, Neth, Jacobs, & Gray, 2013), or as the prior for another Bayesian model. However, choice outcomes and options map nicely onto the Dirichlet distribution concentration parameter and categories respectively. Simply, the categories are effectively predetermined by the number of choices and the probability mass for each category is distributed as a function of the number of rewarding observations.

As an attempt to design a Bayesian analog to the Decay model, the DPD model was extended to include a decay parameter. The Dirichlet Probability Distribution Decay (DPD-Decay) model decays the memory representations of the total number of rewarded outcomes at each timepoint. Critically, this means that additional uncertainty is introduced into the probability distribution. As the memory of rewarded outcomes for each option tends towards 0, all options would have an equiprobable chance of being selected.

While Bayesian models have been criticized in prior research for being simple vote-counting models (Jones & Love, 2011), it's possible that, if each rewarding event is considered a vote, the DPD model will predict similar behavior as the Decay model. This could allow the DPD model to predict a bias toward frequency of reward rather than average reward magnitude, and recent work suggests that reward frequency exerts a larger effect on behavior than average reward magnitude (Worthy, Otto, Cornwall, Don & Davis, 2018). Thus, DPD models with sparse priors may represent a cognitive process of predicting the probability of a rewarding event, based solely on reward frequency. Our goal in the present work is to verify these predictions and examine the degree to which they are consistent with human behavior.

Difference Between Models

The key difference between the Delta model and the reward frequency models (Decay and Dirichlet) is in how each type of model utilizes reward information to make predictions about future choices. The Delta model uses average reward information whereas the Decay and Dirichlet models utilize the frequency of rewards to formulate a cumulative representation of reward. This is important as, per Estes (1976), probability judgements about the choices are heavily influenced by the frequency that each option produces a reward, rather than the average reward value. As such, it would be expected that the when rewarding options are shown in disproportionate frequencies, the predictions of the Delta model and the Decay and Dirichlet models will diverge. It would be expected that tasks which consist of rewards of varying frequency and value would show differences in each models' predictions.

To ascertain the general predictions of each model, and determine the differences therein, three tasks which have previously examined the effect of reward frequency and value were selected to be simulated using each model. To verify the predictions made by each model and task combination, each of the models were fit to human data collected from each of the three tasks.

Experimental Tasks

Iowa Gambling Task. The Iowa Gambling Task (IGT; Bechara, Damasio, Damasio, & Anderson, 1994) allows four options to be chosen from, each with their own reward schedule over the course of 100 total trials. The reward schedule for the IGT can be found in Table 1 below. Traditionally, the task consists of two options which result in a net loss of points, and two options which results in a net gain. Options A and B offer participants larger rewards on gain trials, but also larger losses on loss trials resulting in an overall net loss for both options. In contrast, Options C and D give smaller rewards and losses resulting in an overall net gain for these two options. Within each 10-choice block for each option, the frequency of gains differs between options. Options A and C show infrequent gains relative to Options B and D which are more consistent. Strictly looking at the net positive options, Options C and D should be the favored decks. However, as Bechara et al. observed, there is a preference for choosing Options A and B which have a higher frequency of larger rewards, but results in a net loss of points.

	A		B		C		D	
Trial	**IGT**	**SGT**	**IGT**	**SGT**	**IGT**	**SGT**	**IGT**	**SGT**
1	100	200	100	100	50	-200	50	-100
2	100	200	100	100	50	-200	50	-100
3	-50	200	100	100	0	-200	50	-100
4	100	200	100	100	50	-200	50	-100
5	-200	-1050	100	-650	0	1050	50	650
6	100	200	100	100	50	-200	50	-100
7	-100	200	100	100	0	-200	50	-100
8	100	200	100	100	50	-200	50	-100
9	-150	200	-1250	100	0	-200	50	-100
10	-250	-1050	100	-650	0	1050	-250	650
Net	**-250**	**-500**	**-350**	**-500**	**250**	**500**	**200**	**500**

Table 1: Reward schedules for both the IGT and SGT by Option Letter. This reward schedule is repeated over the total 100 trials.

Soochow Gambling Task. The Soochow Gambling Task (SGT; Chiu et al., 2008) is a task similar in procedure to the IGT aside from a change in the reward schedule of each option. The reward schedule for each option in the SGT can be found in Table 1 below. Similar to the IGT, over the course of 100 trials, participants are able to select one of four options. Options C and D are still the options with an overall net reward gain, and likewise with Options A and B having a net reward loss. Both Options A and B offer participants consistent gains of 200 or 100 points, respectively, followed by a large loss which results in a net loss for both options. Inversely, Options C and D show consistent losses followed by a large gain resulting in a net gain. The gains and losses shown in Options A and B are exactly opposite in terms of sign. Where A and B show consistent rewards followed by a large loss. Importantly, the best options according to overall gain are also the options with the most consistent losses.

Similar to the IGT, there is a large preference for Options A and B indicating that frequency of rewards, despite losses, is a good predictor of choice preference(Byrne & Worthy, 2016).

Binary Choice Task. This task, as presented by Worthy et al. (2018), assesses the effect of reward frequency in a different manner. The task consists of four options, A, B, C, and D, where each have a respective probability of giving a reward of .65, .35, .75, .25. The possible rewards for this task were binary in that the reward totals were either 1 or 0. The task pairs Options A and B, and Options C and D, together and presents them randomly interspersed during training. Importantly, there are 100 AB trials and 50 CD trials which creates a situation where frequency of reward and average reward are in opposition if it is learned that Option A and C are the most rewarding within the respective pairs. The task then consists of 25 transfer trials for each of the remaining pairs of A, B, C, and D and bars further reward feedback. Worthy et al. observed that human participants were more likely to prefer Option A over C on AC pairing trials indicating that despite having a smaller average reward, option A is preferred over option C because of more frequent, and therefore higher cumulative reward.

Method

Model Formalisms

The Delta and Decay rule used in this paper are identical to those described in Worthy et al. (2018). Reward (r) and the expected value (EV) is calculated for each j option on each t trial. The Delta rule is described in Equation 1 as:

$$EV_{j,t+1} = EV_{j,t} + \alpha \cdot (r_t - EV_{j,t}) \cdot I_j \qquad (1)$$

Where I_j is a variable which indicates option choice via a value of 1 if j option is chosen on trial t, and 0 otherwise. This formulation ensures that only the expected value for the chosen option is updated, and the other options, whether seen or not, are not updated. Alpha (α) is denoted as a learning rate parameter where $\alpha \in (0,1)$. For the Delta model in particular, α modifies the $(r_t - EV_{j,t})$ prediction error by giving greater weight to more recent outcomes with higher α values, and lower α values giving less weight to recent outcomes and producing little change in the expected value on each trial.

Similarly, the Decay model tracks changes in expected value, but instead of updating the expected value by way of a prediction error the raw reward value is used. However, this does not mean that expected value consistently increases for each chosen option. On each trial, each j option will be modified by a decay parameter (A; $A \in (0,1)$) regardless of whether the j option was seen or chosen. Critically, this means that the expected value for each option will decay over time and only increase when a reward for that option is received. Thus, the more frequent the reward, the greater the expected value. The formula for computing the change in expected is described below in Equation 2:

$$EV_{j,t+1} = EV_{j,t} \cdot A + r_t \cdot I_j \qquad (2)$$

As mentioned above, the DPD model focuses solely on the number of times each j option is rewarded (r) and uses that information to update a Dirichlet probability distribution. Simply, a Dirichlet distribution takes k, the total number of j options, and their respective number of rewarded trials (γ_j) and produces a probability density (x_j) for each j option where $x_j \in (0,1)$ and $\sum_{j=1}^{k} x_j = 1$. In other words, the updating of the distribution occurs in two steps as described in Equations 3 and 4:

$$\gamma_{j,t+1} = \gamma_{j,t} + r_t \cdot I_j \qquad (3)$$

$$f(x_{1,t+1} \ldots x_{k,t+1} | \gamma_{1,t+1} \ldots \gamma_{k,t+1}) = \frac{1}{B(\gamma)} \prod_{j=1}^{k} {x_{j,t}}^{\gamma_{j,t}-1} \qquad (4)$$

$$\text{where } B(\gamma) = \frac{\prod_{j=1}^{k} \Gamma(\gamma_j)}{\Gamma(\sum_{j=1}^{k} \gamma_j)}$$

On each t trial, the reward value for one option is added to the chosen option which will distribute slightly more probability density to the chosen option. To determine choice with this model, a random sample is taken from the Dirichlet distribution which results in a simplex, or a vector of probabilities which sum to 1. Critically, this implies that as one option is rewarded more frequently, the probability value sampled from the distribution will tend to be of greater value, and thus the option is more likely to be chosen. Taking a single sample, rather than integrating over the posterior, was a decision made with the assumption that this would better reflect human performance as the beliefs surrounding each option is uncertain. As more information is learned about an individual option, the belief about the positive outcomes of that option will become more certain, and thus the probability of choosing that outcome will be more consistent.

An extension of the DPD model presented above, the DPD-Decay model includes the decay parameter (A) which decays the total number of rewarded trials (γ_j) for each option on each trial similar to how the Decay model functions. By decaying the rewarded trial values, the model increases the amount of uncertainty and allows a greater range of possible values to be randomly sampled. This also implies that the more frequently an option is seen the more likely it is to overcome the consistent decay, such that it is granted more probability density over time. Expressly, the decay parameter in this equation will weigh the model for or against more recent outcomes. In Equation 5, $\gamma_{j,t+1}$ is computed for every j option and are subsequently inserted into Equation 4.

$$\gamma_{j,t+1} = \gamma_{j,t} \cdot A + r_t \cdot I_j \qquad (5)$$

For the Delta and Decay models, the predicted probability that any given option j is chosen C on a particular trial t, P($C_{j,t}$), is calculated by way of a Softmax choice function shown in Equation 6 below:

$$P|C_{j,t}| = \frac{e^{\beta \cdot EV_{j,t}}}{\sum_{1}^{N(j)} e^{\beta \cdot EV_{j,t}}} \qquad (6)$$

Like the Yechiam & Ert (2007) Softmax application used in Worthy et al. (2018), $\beta = 3^c - 1; c \in (0,5)$, where c is an inverse temperature parameter which dictates how often the option with the higher expected value is chosen. When c approaches 0, choices are more random. Inversely, choices are weighted more heavily towards the options with the highest expected value as c approaches 1. Simply, this choice function determines the probability of choice by computing the proportion of the scaled chosen option divided by the sum of the scaled choice and alternative choices.

Simulation and Behavioral Methods

For each task, 10000 simulated participant datasets were created with randomized model parameters of α, A, and c, for applicable models, for each participant. Each of these parameters were drawn from a uniform distribution: $U(0,1)$ for learning and decay rates, and $U(0,5)$ for the inverse temperature parameter. These parameters were kept consistent across models within each simulation, but each model ran independently in regard to the choices made and corresponding output. The output for each of these simulations was the probability of choosing each outcome, the expected value of each option, and the choices made on each trial.

For each task, human behavioral data was collected from an undergraduate population with sample sizes of ~50 for each task. Each participant completed the experiment in a Psychtoolbox 2.54 environment on a Windows computer running Matlab. The general procedures used in the simulations were identical to the computerized version of the tasks that participants completed, however graphical and counterbalancing considerations were needed for real participants that are detailed below for each experiment.

In both the IGT and SGT, the options were displayed onscreen as a deck of cards, each with their own random color. The onscreen location of each individual deck was displayed from left to right in a random arrangement of Options A-D for each participant. Upon selecting a deck, the participant would be shown the card being overturned and the amount of reward. Additionally, participants were given a set amount of points in an onscreen bank that would increase or decrease depending on the outcome.

For the Binary Choice Task, each of the four options were randomly assigned a fractal image randomly drawn from a pool of 12 images. Like the IGT and SGT, the order of the 4 selected images were randomly arranged on screen from left to right. However, Options AB and CD were always together as a pair, but the order of each pair varied for each participant. As an example, some potential orderings of the option could include: ABCD, CDAB, BACD, etc. Each selection of an option showed the option turning over to reveal the outcome of that trial. Importantly, and consistent with the simulations, reward feedback only occurred during the initial 150-trial training phase, but the transfer phase, participants were only shown a gray outline around the option they chose instead of the point value they would have seen on the training trials.

Results

Simulation Output

For the IGT and SGT, the simulation metric that will be reported is the overall performance on the task as computed by subtracting the sum of the net loss options from the sum of the net gain options: (A+B)-(C+D). For both the IGT and SGT, the performance of each model, and the actual participant data for comparison, is plotted over all 100 trials in Figures 1 and 2. In the IGT the Delta model was more likely to choose the net gain options over the more frequently rewarding net loss options. The Decay model also showed a preference for the net gain options overall. Both the DPD and DPD-Decay models showed no preference for either the net gain or loss options, but this behavior also seems to be reflected, albeit slightly, by the actual participant data which rapidly varies in preference for either the net gain or loss options over time.

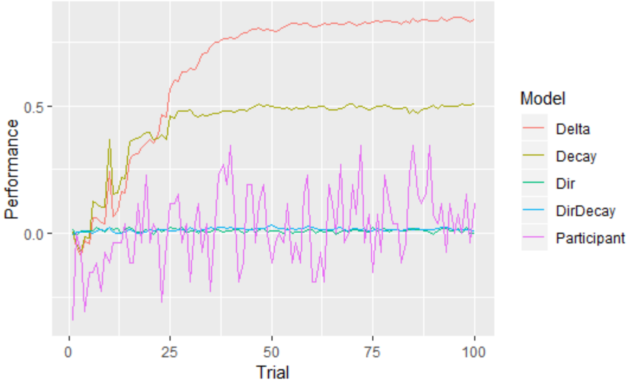

Figure 1: Average performance on the IGT by model and actual participant data.

In the SGT, the Delta model again showed a preference for the net gain options, but the Decay model now shows behavior that greatly reflects the behavior shown by actual participants. Both the human and simulated Decay model datasets showed an initial preference for the net loss options, but over time began to tend towards the net gain options which is consistent with prior research as previously discussed. The DPD and DPD-Decay model again showed similar results, but in the SGT, they show a large preference for the more frequently rewarding net loss options.

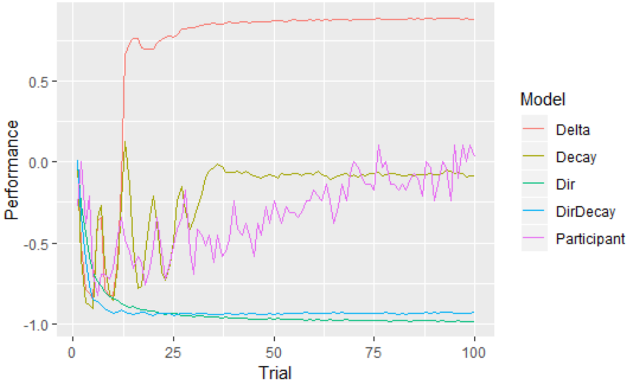

Figure 2: Average performance on the SGT by model and actual participant data.

For the Binary Choice Task, as shown in Figure 3A-B, each model was able to learn that there is a more rewarding option in each option pair. However, the rate at which the most rewarding, or best, option was identified and overall preference for the best option differed between models. The Delta model showed the greatest preference for the best options out of the four models, followed by the DPD, Decay, and DPD-Decay. When solely learning which option has the largest average reward, it is no surprise that the Delta model outperforms the other three models. However, when looking at the choice predictions for the remaining option pairs, as shown in Figure 3C, a difference between the models emerge. The Delta model predicts more C choices, whereas the Decay, DPD, and DPD-Decay models all predict more A choices. The remaining option pairs showed relatively similar predictions since there was not as big of a discrepancy between an options' expected value and number of observations. The large peaks in the DPD model are indicative of the frequency of outcome observations for each pair. The more outcomes observed, the more likely the model will choose the same option.

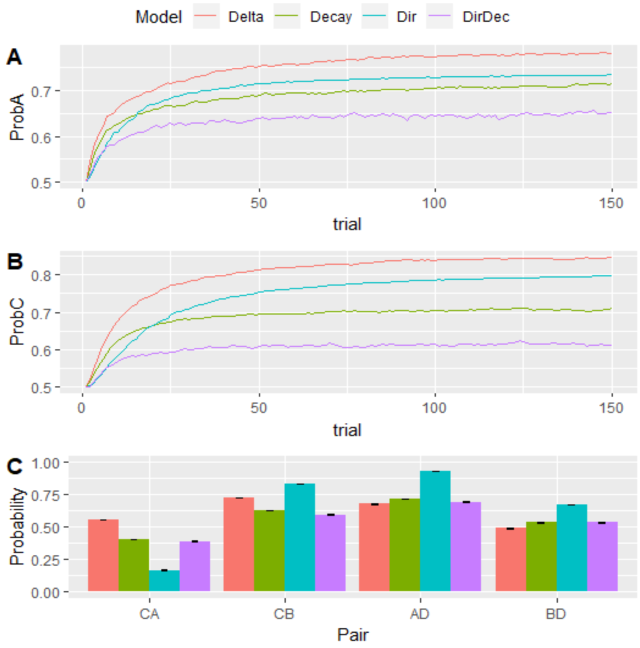

Figure 3: A and B show the probability of choosing the best option, either Option A or C respectively, over the course of 150 trials. C shows the predicted probability of choosing the best options if the simulated participant were to see the remaining pairings of options.

Behavioral Fits and Comparisons

Participants were independently recruited for one of the three tasks from an undergraduate sample. Each participant was reimbursed for their time with partial completion of course credit. For each task we recruited comparable sample sizes: 52 participants for the IGT; 58 participants for the SGT; 50 participants for the Binary Choice Task.

Each of the models were directly fit to the behavioral data by maximizing the likelihood of each model via the 'optim' function with a 'L-BFGS-B' method in R. The decay and inverse temperature parameters were included as free parameters for the respective models. The Delta and Decay models utilized two free parameters while the DPD-Decay model used only the decay parameter. No free parameters were used in the DPD model for the IGT and SGT. When fitting the Binary Choice Task data alone however, the DPD and DPD-Decay model included an inverse temperature parameter. In this task, for both models, the probability simplex was drawn from the Dirichlet distribution, as previously discussed, but the values relevant to the two observed options were used in the softmax function to compute a choice probability for each option which summed to 1.

A Bayesian Information Criterion (BIC; Schwarz, 1978) value was computed for each individual participant within each model and used to calculate the average BIC and subsequent BIC differences between each model. BIC was calculated by calculating the deviance of the model and adding additional error based on the number of free parameters k and number of trials t: $-2ln(L) + (k \cdot ln(t))$. Lower BIC values indicate a better fit to the behavioral data. As per Wagenmakers (2007), the BIC difference between the models can additionally be used to calculate a Bayes Factor which would show evidence for one model over another: $\mathrm{BF}_{10,\mathrm{Model1}}=\exp((\mathrm{BIC}_{\mathrm{model2}}-\mathrm{BIC}_{\mathrm{model1}})/2)$.

Table 2 below details the BIC values of each model for each task along with the best fitting parameters for each model. For the IGT and Binary Choice Task, the Decay model shows an advantage over the other models. For the IGT, the next best fitting model was the DPD with a BIC of 268.9 which is shown to be significantly different from the Decay model with a Bayes Factor (BF) of 3.33. BFs with values greater than 3, or less than 1/3, are believed to have adequate evidence to reject the null hypothesis that the models are equal. The Decay model in the SGT was the next best fitting model behind the DPD-Decay model with BIC values of 269.8 and 267.8 respectively. This difference, with a BF of 2.7416, shows that both models are similar in their fits of the SGT data. In the Binary Choice Task, the Decay model BIC (279.7) is closely followed by both the DPD and DPD-Decay models; 282.8 and 280.5 respectively. The difference between the Decay and DPD model is significant with a BF of 5.1984, but there is not enough evidence to say that the Decay and DPD-Decay models are different, BF = 1.5115.

Table 2: Average Model Values

		Best a or A	Best c	BIC
IGT	Delta	.1009	.3756	278.0677
	Decay	**.6857**	**.00538**	**266.4658**
	DPD	N/A	N/A	268.8740
	DPD-D	.0218	N/A	273.8372
SGT	Delta	.4613	.3564	274.4863
	Decay	.5268	.0019	269.7714
	DPD	N/A	N/A	282.9299
	DPD-D	**.1454**	**N/A**	**267.7543**
Binary	Delta	0.3821	1.5120	296.2498
	Decay	**0.1765**	**0.4978**	**279.7178**
	DPD	N/A	1.3088	282.8315
	DPD-D	0.8770	1.5673	280.5440

It was also of interest to determine the proportion of participants whose data were best fit by each model. To do this, each non-redundant combination of models for each task was examined to figure how many participants' data were best fit by each model. Table 3 presents the proportion value for each model combination by task. The first model listed in the pair is the reference model. Values shown in bold represent that the reference model was the best fitting model of the pair.

As expected from the data in Table 2 for the IGT, the Decay and DPD model best fit the largest proportions of participants, and the DPD model showed the best fit overall. For the SGT, despite showing a large average BIC value, the Delta model showed a better fit slightly more participants than the Decay model, but not the DPD-Decay model. Additionally, the Decay model, rather than the DPD-Decay model, was the best fitting model for most participants. In the Binary Choice Task, the DPD models better fit more participants than the Decay model which showed the better fit on average. The DPD model showed the overall highest proportion best fit on this task as well.

Table 3: Proportion Best Fit

By Model	IGT	SGT	Binary
Delta<Decay	.44	**.54**	.38
Delta<DPD	.06	**.54**	.30
Delta<DPD-D	.44	.48	.38
Decay<DPD	.33	**.66**	.40
Decay<DPD-D	**.71**	**.62**	.44
DPD<DPD-D	**.83**	.47	**.70**
Overall	**IGT**	**SGT**	**Binary**
Delta	.05	.24	.26
Decay	.24	**.34**	.16
DPD	**.53**	.22	**.44**
DPD-D	.07	.19	.14

Discussion

The simulations and experiment presented in this paper examined the influence of reward frequency and probability on choices made in a decision-making task. Four models were compared that made both convergent and divergent predictions about which option was more valuable in three tasks which examine the effect of reward frequency. As similarly described by Worthy et al. (2018), there were divergent simulation predictions between the Delta rule and the reward frequency models where the Delta rule more often chose the options with the higher value rewards, whereas the reward frequency models, the Decay, DPD, and DPD-Decay, tended to choose the options which resulted in the most frequent rewards. The data from the experimental tasks showed that human participants more often chose the more frequent options in most cases. This behavior is in support of the predictions of all three of the reward frequency models. This is shown in which models where the best fitting model on average. For all three tasks, the best fitting model was a model which attended more towards the frequency of reward rather than the average value of reward. However, there also seems to be some individual differences in people who attend more towards average reward value instead of the frequency of reward. This can best be seen when looking at the SGT and Binary Choice Task. For both of these tasks, there was a sizable subset of participants who were best fit by the Delta model than the other three models.

There also exists some important differences in the reward frequency models despite their similarities. One of which is between the DPD and DPD-Decay models and the Decay model. When looking at Figure 1, the performance values for the DPD and DPD-Decay model are fairly constant about 0. This is most likely due to how the Dirichlet models compute reward. These models do not consider reward value, only the observation of a reward. Looking back to the reward schedules for the IGT, one net gain and one net loss option have fairly frequent rewards. With how the performance calculation considers the number of choices, and how the Dirichlet models determine choice by the number of observed rewards, you can begin to see how the number of net gain and loss choices would be about equal, and thus result in a performance of ~0. This can also be seen in the simulation of the SGT as well in Figure 2. The two Dirichlet models show an overwhelming preference for the net loss options. Again, looking at the reward schedule, the net loss options are the only options that have a frequent occurrence of reward as the net gain decks only give a reward every 5 successive picks. These two Dirichlet models may aid in making sense of the "Deck B" phenomenon in the SGT where people tend to choose the net loss options since the reward most frequently. However, the average fit for the DPD model was quite large. Which suggests that pure frequency of reward is not entirely predictive of choice on the SGT. With the DPD-Decay model showing the best average fit, this suggest that the frequency of reward is predictive, but that the overall representation of the total number of rewarded outcomes decays over time.

For the DPD model in particular, another difference be seen in Figure3C with the large peaks in the option pair predictions relative to the other models. Like detailed for the IGT and SGT, these peaks can be explained by looking at the rate of reward and frequency of observing the option pair. For these option pairs the best option is the one that is either the most frequently seen and/or rewarded. Thus, the model would be more likely to choose these options.

However, this also ties in to the major conclusion of this paper, that despite not utilizing any reward information, these Dirichlet models are able to fit human behavioral data on three tasks relatively well solely using a count of rewarding outcomes. Generally, choice selection may depend on reward value when all other factors are equal, but if rate of reward changes or if there is knowledge of number of previously rewarding outcomes, frequency of reward may take precedence over reward value. Though, like shown by the proportion of best fitting models, there may be a subset of people who focus on the overall reward value regardless of the frequency of the outcomes.

References

Annis, J., & Palmeri, T. J. (2018). Bayesian statistical approaches to evaluating cognitive models. *Wiley Interdisciplinary Reviews: Cognitive Science*, *9*(2)

Busemeyer, J.R., & Stout, J.C. (2002). A contribution of cognitive decision model to clinical assessment: Decomposing performance on the Bechara Gambling Task. *Psychological Assessment, 14,* 253-262.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1–3), 7–15.

Byrne, K. A., & Worthy, D. A. (2016). Toward a mechanistic account of gender differences in reward-based decision-making. *Journal of Neuroscience, Psychology, and Economics*,*9*(3–4),157–168.

Chiu, Y.-C., Lin, C.-H., Huang, J.-T., Lin, S., Lee, P.-L., & Hsieh, J.-C. (2008). Immediate gain is long-term loss: Are there foresighted decision makers in the Iowa Gambling Task? *Behavioral and Brain Functions*, *4*(1), 13. https://doi.org/10.1186/1744-9081-4-13

Daw, N., O'Doherty, J., Dayan, P., Seymour, B., & Dolan, R. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441,* 876-879.

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.

Estes, W.K. (1976). The cognitive side of probability learning. *Psychological Review, 83,* 37-64.

Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*(1), 1-12.

Gluck, M.A., & Bower, G.H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 128,* 309-331.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition.

Griffiths, T. L., Sanborn, A. N., Canini, K. R., & Navarro, D. J. (2008). Categorization as nonparametric Bayesian density estimation. *The probabilistic mind: Prospects for Bayesian cognitive science*, 303-328.

Jacobs, R.A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks, 1,* 295-307.

Jacobs, R. A., & Kruschke, J. K. (2011). Bayesian learning theory applied to human cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(1), 8-21.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169-188.

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.

Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using Dirichlet processes. *Journal of mathematical Psychology*, *50*(2), 101-122.

Otto, A. R., & Love, B. C. (2010). You don't want to know what you're missing: When information about forgone rewards impedes dynamic decision making. *Judgment and Decision Making, 5*(1), 1–10.

Pang, B., Blanco, N. J., Maddox, W. T., & Worthy, D. A. (2017). To not settle for small losses: evidence for an ecological aspiration level of zero in dynamic decision-making. *Psychonomic bulletin & review*, *24*(2), 536-546.

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.

Rumelhart, D.E., McClelland, J.E. & the PDP Research Group. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* vols. 1 and 2. Cambridge, MA: MIT Press.

Sims, C. R., Neth, H., Jacobs, R. A., & Gray, W. D. (2013). Melioration as rational choice: Sequential decision making in uncertain environments. *Psychological Review*, *120*(1), 139.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

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

Widrow, B., & Hoff, M.E. (1960). Adaptive switching circuits. *1960 WESCON Convention Record Part IV,* 96104.

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

Worthy, D. A., Otto, A. R., Cornwall, A. C., Don, H. J., & Davis, T. A Case of Divergent Predictions Made by Delta and Decay Rule Learning Models. *Proceedings of the Cognitive Science Society.*

Yechiam, E., & Busemeyer, J.R. (2005). Comparison of basic assumptions embedded in learning models for experiencebased decision-making. *Psychonomic Bulletin & Review, 12,* 387-402.

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