

What cognitive processes drive response biases? A diffusion model analysis

Fábio P. Leite^{a*} and Roger Ratcliff^{b†}

Department of Psychology

^a The Ohio State University, Lima, OH

^b The Ohio State University, Columbus, OH

Abstract

We used a diffusion model to examine the effects of response-bias manipulations on response time (RT) and accuracy data collected in two experiments involving a two-choice decision making task. We asked 18 subjects to respond “low” or “high” to the number of asterisks in a 10×10 grid, based on an experimenter-determined decision cutoff. In the model, evidence is accumulated until either a “low” or “high” decision criterion is reached, and this, in turn, initiates a response. We performed two experiments with four experimental conditions. In conditions 1 and 2, the decision cutoff between low and high judgments was fixed at 50. In condition 1, we manipulated the frequency with which low- and high-stimuli were presented. In condition 2, we used payoff structures that mimicked the frequency manipulation. We found that manipulating stimulus frequency resulted in a larger effect on RT and accuracy than did manipulating payoff structure. In the model, we found that manipulating stimulus frequency produced greater changes in the starting point of the evidence accumulation process than did manipulating payoff structure. In conditions 3 and 4, we set the decision cutoff at 40, 50, or 60 (Experiment 1) and at 45 or 55 (Experiment 2). In condition 3, there was an equal number of low- and high-stimuli, whereas in condition 4 there were unequal proportions of low- and high-stimuli. The model analyses showed that starting-point changes accounted for biases produced by changes in stimulus proportions, whereas evidence biases accounted for changes in the decision cutoff.

Keywords: Bias manipulation; diffusion model; stimulus frequency; payoff; drift criterion.

1 Introduction

It is well known that experimental manipulations in perceptual decision making tasks produce systematic changes in the behavioral responses of the subjects. Experimental psychologists, for example, have observed that subjects bias their responses based on the probability of occurrence of the stimulus (Falmagne, 1965; Jarvik, 1951; Kirby, 1976; Laming, 1969; for a review, see Luce, 1986).¹ Formal models have been used to describe and predict response bias in perceptual two-choice tasks (e.g., Edwards, 1965; Gold & Shadlen, 2000; Ratcliff, 1985). These models typically conceptualize the de-

cision process as the accumulation of sensory information over time towards a decision threshold. The aim of this study was to use one of these models (viz., Ratcliff’s diffusion model, 1978) to examine what processes account for biases in two-choice tasks produced by manipulations of stimulus frequency, payoffs, and movement in the decision cutoff (i.e., the point at which stimuli are assigned to one versus the other response category).

In manipulations of stimulus frequency, it is well known that response times (RTs) to stimuli that appear more often are faster than RTs to stimuli shown less often. In addition, accuracy is increased from less to more frequent stimuli. Remington (1969), for example, used a task in which subjects responded to one of two lights by depressing corresponding keys. There were blocks of trials in which each light was turned on in half of the trials, blocks in which one light was turned on in 70% of the trials and the other in 30% of the trials, and blocks in which these proportions were reversed. In comparison to the equally likely condition, response times (RTs) were faster for trials in which a light appeared more often and slower for trials in which it appeared less often.

*Correspondence concerning this article should be addressed to Fábio P. Leite, Department of Psychology, The Ohio State University, Lima, Ohio 45804. E-mail: leite.11@osu.edu.

[†]This study was partially supported by National Institute on Aging grant R01-AG17083 and by Air Force grant FA9550-11-1-0130, awarded to Roger Ratcliff. We thank Jon Baron, Adele Diederich, and an anonymous reviewer for their comments on this manuscript.

¹Because we studied a perceptual task, we refer to response bias to describe response patterns that demonstrate subjects favoring or avoiding one of the alternatives, not connoting that the response was made in error or as a result of some error in judgment.

Studying the effects of explicit payoff manipulations has also been of long-standing interest in cognitive psychology (cf., Edwards, 1965, p. 322). Fitts (1966), for example, used payoff matrices to allow subjects to earn a bonus at the end of a session. Subjects always earned one point for each correct and fast response, but lost points based on one of the following structures: One, they lost half a point for correct and slow responses, one tenth of a point for wrong and fast and responses, or one point for wrong and slow responses; Two, they lost one tenth of a point for correct and slow responses, half a point for wrong and fast responses, or one point for wrong and slow responses. Fitts found that subjects either responded faster and made more errors or responded more slowly and made fewer errors, in order not to lose too many points. More recently, Rorie, Gao, McClelland, and Newsome (2010) manipulated response biases with payoffs in a motion discrimination task with multiple reward contingencies and four reward conditions. The two rhesus monkeys tested showed indistinguishable psychophysical performance in the neutral reward conditions but biased performance (i.e., faster response and increased response probability) toward the high-reward response in biased conditions.

Our present study manipulated stimulus frequency and payoffs as typically reported in the literature, but planned a novel direct comparison of the two manipulations in an attempt to answer whether the effect observed on both RT and accuracy due to changes in stimulus frequency is similar to the effect due to changes in reward values. Moreover, we added to this comparison a less commonly studied manipulation of decision cutoffs. Because we compared these response-bias manipulations with the aid of the diffusion model, we present relevant information on computational models below.

1.1 Computational models

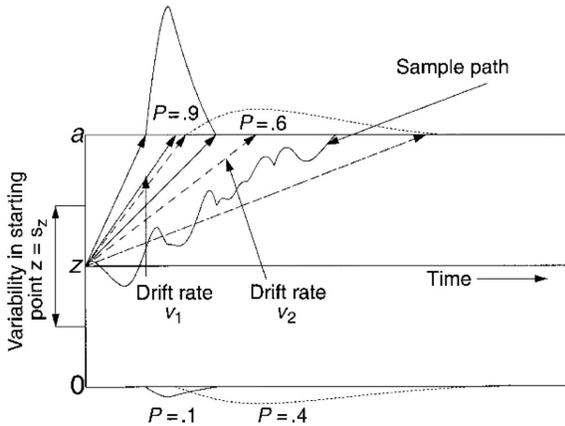
Commonly, computational models have been applied to experimental data to infer the properties of the underlying sensory representation and processing leading to decisions. In the study of perceptual decision making in particular, a computational model that has allowed psychologists to infer properties of the underlying sensory representation from behavioral evidence is signal detection theory (SDT; Swets, Tanner Jr., & Birdsall, 1961; Green & Swets, 1966). In a typical paradigm to which SDT is applied, subjects must detect the presence of a signal (shown in a proportion of trials) embedded in noise (always present). It is assumed that subjects set a threshold criterion that the amplitude of the signal must surpass so that a “signal” is reported. It is further assumed that this criterion setting is influenced by the prior probability of the signal. In the model, the parameter associated

with the threshold criterion measures the response bias produced by the criterion setting. Analogously, in perceptual judgments involving two-choice paradigms, it is often assumed that the observer will set a criterion between the two alternatives. When the observer judges the information obtained from the stimulus to be above this criterion, then one choice is reported; otherwise, the other choice is reported.

Whereas SDT can account for accuracy, it cannot account for RT (e.g., Ratcliff, Van Zandt, & McKoon, 1999, Figure 15). To account for accuracy and RT at the same time, researchers have turned to sequential sampling models, among other families of models, to fit both dependent variables simultaneously. One sequential sampling model is the Wiener diffusion process model (Ratcliff, 1978)—herein the diffusion model. In the diffusion model, stimulus evidence is assumed to be accumulated gradually over time, from a starting point (z) toward one of two alternative decision boundaries (Figure 1.1). The separation between these two boundaries, or the amount of evidence required for a response, is modeled by parameter a . The more evidence present in a stimulus, the faster the accumulation process reaches a boundary, at which point the accumulation is terminated and a response is made. The rate at which information is accumulated, the drift rate (ν), is a function of the quality of stimulus information. Variability within a trial in the accumulation of information (noise) gives rise to variability in RTs and allows the process to hit the wrong boundary, giving rise to errors. Across-trial variability in the values of drift rate (η) and starting point (s_z) is required because it is assumed subjects cannot accurately set these parameters to have the same value from trial to trial (e.g., Laming, 1968; Ratcliff, 1978).

In addition to the decision components of processing, all other components (such as stimulus encoding and response execution) are combined into one nondecision-time parameter and have a mean duration of T_{er} . This nondecision time is assumed to vary across trials, with values coming from a rectangular distribution over the range s_t . Consequently, the predicted mean RT is the mean time for the decision process to terminate plus the nondecision time governed by T_{er} and s_t . As noted in Ratcliff and Tuerlinckx (2002, p. 441) and in Ratcliff and Smith (2004, p. 338), the assumption about the shape of the nondecision time distribution is not crucial because the standard deviation of the distribution of decision times is four or more times larger than that of the distribution of nondecision times. As a result, the shape of the RT distribution is determined almost completely by the shape of the distribution of decision times. Variability in the nondecision components, which was crucial to a successful diffusion model account of lexical decision data in Ratcliff, Gómez, and McKoon (2004), is included

Figure 1: Illustration of the decision process of the diffusion model. Two mean drift rates, ν_1 and ν_2 —subject to across-trial variability (η), represent high and low rates of accumulations of evidence. Accumulation of evidence in each trial starts at point z , subject to across-trial variability (s_z). The accumulation process terminates after it crosses either boundary (a or 0). Correct responses are made when the accumulation process crosses a , whereas incorrect responses are made when it crosses 0 . The three solid-line trajectories illustrate fast processes around ν_1 , and the three dashed-line trajectories illustrate slow processes around ν_2 . In combination, they show how equal steps in drift rate map into skewed RT distributions. Predicted mean RT is the mean time for the decision process to terminate plus a nondecision time (including processes such as stimulus encoding and response execution) governed by T_{er} , subject to across-trial variability (s_t).



because it has two effects on model predictions: the leading edge of the RT distribution (i.e., the point at which 10% of the responses had terminated) has greater variability across conditions with it than without it; and the rise in the leading edge of the RT distribution is more gradual with it than without it.

In the diffusion model, there are two ways to model response bias (Ratcliff, 1985; Ratcliff et al., 1999; Ratcliff & McKoon, 2008). One, analogous to the change in criterion threshold in SDT, involves a shift in the criterion that separates positive from negative drift rates (Figure 1.1). That is, one parameter (referred to as a “drift criterion” parameter, “ d_c ”) produces an estimate of the amount to be added to or subtracted from the mean drift rates between conditions such that a null mean drift rate (a horizontal line starting at z in Figure 1.1) in an unbiased condition would be shifted to a positive or negative mean drift rate in a biased condition, for example. For small to moderate biases in drift rates, there are only small changes in the leading edges of the RT distributions between biased and

Figure 2: Illustration of the drift criterion explanation of the effects of response probability manipulations on response bias in the diffusion model. When the probability of response A is higher, the drift rates are ν_a and ν_b , with the zero point close to ν_b . When the probability of response B is higher, the drift rates are ν_c and ν_d , and the zero point is closer to ν_c (cf. Ratcliff & McKoon, 2008, Figure 3, bottom panel).

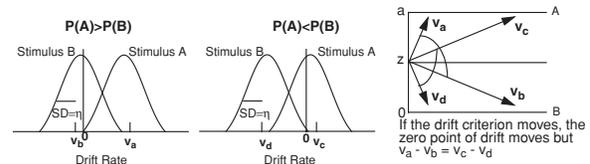


Figure 3: Illustration of the starting point explanation of the effects of response probability manipulations on response bias in the diffusion model. When A and B are equally likely, the process of accumulation of evidence starts equidistantly from boundaries 0 and a . When the probability of response A is higher, the starting point is closer to a than to 0 . When the probability of response B is higher, the starting point is closer to 0 than to a (cf. Ratcliff & McKoon, 2008, Figure 3, top panel).



unbiased conditions. The other way to model response bias involves moving the starting point of the diffusion process nearer the boundary toward which the responses are biased (Figure 1.1). This predicts a shorter leading edge of the RT distribution in the biased condition than in the unbiased condition (Ratcliff, 1985, 2002).

Although these two methods of modeling response bias and the experimental manipulations that produce response bias are known, there has been no study to compare the effects of the different response-bias manipulations we present here. Below we summarize our objectives and choices of design and analysis.

1.2 Study overview

Our aim with this article was to produce a comprehensive model-based account of response-bias manipulations. To that end, we chose to test four tasks (stimulus frequency, payoff structure, decision cutoff, and stimulus frequency with decision cutoff) in the same two-choice numerosity discrimination and to apply the diffusion model to the data for the analysis of response biases in each of the

tasks. The key questions of our study were: Do manipulations of stimulus frequency and payoff structure affect the same component parameters driving the decision and in the same way? How does the manipulation of stimulus frequency compare to that of decision cutoff? What happens when stimulus frequency and decision cutoff are manipulated at the same time? Answers to these questions will contribute to the current literature on perceptual decision making, in both cognitive and cognitive neuroscience domains, by informing which cognitive processes are more strongly (or similarly) associated with each type of task.

We chose a payoff-structure manipulation that involved mapping response alternatives to reward values (i.e., points to be converted into small financial incentives). In a two-choice task, subjects typically favor one response alternative over the other because of a high reward (or a low cost) value associated with the one or because of a high cost (or a low reward) value associated with the other response. In the emerging field of neuroeconomics, this type of manipulation is an example of a value-based decision (for reviews, Glimcher, 2005; Gold & Shadlen, 2007; Sanfey, Loewenstein, McClure, & Cohen, 2006; Sugrue, Corrado, & Newsome, 2005).

Our choice of a model analysis allowed us to test the following hypotheses involving cognitive processes: stimulus frequency and payoff structure affect the same component parameters driving the decision in the same way (H1); stimulus frequency affects starting point, whereas decision cutoff affects drift criterion (H2); and, when stimulus frequency and decision cutoff are manipulated at the same time, both starting point and drift criterion are affected (H3). Testing these different hypotheses formally is in line with neuroeconomics' goal of uniting formal models of subjective variables, descriptions of the nature of cognitive processes, and neural events by means of computational models capable of identifying the signals and signal dynamics that are required by different problems (Rangel, Camerer, & Montague, 2008).

Our choice of the diffusion model was due to its ability to fit RT and accuracy data from several two-choice tasks successfully, while allowing for sensible parameter interpretation across conditions. Ratcliff (2002), for example, related RT and accuracy in a brightness discrimination task in which accuracy varied across a relatively wide range. In fits of the diffusion model, drift rate increased as a function of stimulus brightness. Furthermore, the diffusion model framework has also been applied recently to value-based choices (Krajbich, Armel, & Rangel, 2010; Milosavljevic, Malmaud, Huth, Koch, & Rangel, 2010). Nevertheless, other sequential sampling models or other diffusion model variants could have been chosen (e.g., Usher & McClelland, 2001). Diederich and Busemeyer (2006) and Diederich (2008), for example, investigated

how sequential sampling models might represent the effects of payoffs, finding that a two-stage accumulation process provided a successful account of the data, suggesting that individuals may initially process the payoff information, then switch to the stimulus information.

In what follows, we report two experiments based on a simple two-choice perceptual task. We start with the general method employed in both experiments, followed by specific sections to each experiment in which we present the results. In Experiment 1, we found that starting point in the model was the crucial parameter to account for changes in stimulus frequency, that changes in payoffs affected starting points less than changes in stimulus frequency, and that shifts in drift rates in the model were crucial to account for changes in decision cutoff. In Experiment 2, we replicated these findings in a within-subjects manipulation, supporting that different cognitive processes are likely driving the different response biases.

2 General method

2.1 Task overview

We chose a two-choice numerosity discrimination in which we asked the subjects to respond "low" or "high" to the number of asterisks in a 10×10 grid, based on a experimenter-determined decision cutoff. We chose this numerosity discrimination because it is a simple task for subjects to perform and for experimenters to manipulate both stimulus frequency and decision cutoff and to incorporate payoffs. In addition, there are few perceptual or memory limitations to the task, to which the diffusion model has been successfully applied previously (Geddes, Ratcliff, Allerhand, Childers, Frier, & Deary, 2010; Ratcliff, 2008; Ratcliff, Love, Thompson, & Opfer, in press; Ratcliff, Thapar, & McKoon, 2001, 2010; Ratcliff & Van Dongen, 2009; Ratcliff et al., 1999). Despite its simplicity, however, this task may relate to practical real-world abilities. Halberda, Mazocco, and Feigenson (2008), for example, found correlates to mathematical achievement in a similar numerosity-discrimination task.

2.2 Stimuli and procedure

The experiment was run on personal computers running Linux operating system with a customized real-time system. Computers were connected to a 17" monitor with resolution of 640×480 pixels and a standard 102-key keyboard.

A computer program was designed to present the instructions, run the trials, and record RT and accuracy. Subjects were instructed to reach a decision regarding the numerosity of asterisks on the screen (i.e., categorized as a low number or as a high number) as quickly

and as accurately as possible. The stimulus display was constructed by randomly placing a number of asterisks on a 10×10 grid (with no border lines) on the upper-left corner of the screen. The difficulty of each stimulus was a function of how many asterisks were present and the cutoff number that separated low-number from high-number responses. For example, a stimulus containing 49 asterisks was to be categorized as a low-number stimulus when the cutoff number (referred to as decision cutoff) was 50. In Experiment 1, the number of asterisks ranged from 31 to 70, whereas in Experiment 2 it ranged from 36 to 65.

The stimulus was displayed until the subject responded, by either pressing “Z” for low-number responses or “/” for high-number responses. RT was measured from stimulus onset to response. Accuracy feedback was provided by displaying “ERROR” after an incorrect response (for 500 ms in Experiment 1 and for 400 ms in Experiment 2). If a response was produced too quickly (likely the result of a guess), feedback was presented using a “TOO FAST” message, displayed (for 500 ms in Experiment 1 and for 400 ms in Experiment 2) after responses faster than 250 ms (in pilot tests, few correct responses were reached in less than 250 ms) in Experiment 1 and faster than 100 ms in Experiment 2. If a response was produced too slowly (possibly the result of a failed attempt to respond or of a missed signal), feedback was presented using a “Too slow” message, displayed for 400 ms after responses slower than 1300 ms in Experiment 2. The next trial started 400 ms and 250 ms after these post-response messages for Experiments 1 and 2, respectively.

2.3 Design

The four experimental manipulations of task structure that follow were used in both experiments. Conditions 1 and 2 were compared to test whether stimulus frequency and payoff structure affected the same component parameters driving the decision in the same way (H1). Conditions 1 and 3 were compared to test whether stimulus frequency affected starting point and decision cutoff affected drift criterion (H2). Conditions 1, 3, and 4 were compared to test whether stimulus frequency and decision cutoff, when manipulated simultaneously, affected both starting point and drift criterion (H3).

- Condition 1: Stimulus frequency.

We set the decision cutoff at 50 such that the subjects were asked to decide whether each stimulus contained a low number of asterisks (< 51) or a high number of asterisks (> 50). We manipulated the proportion of low to high stimuli such that there were unbiased (equal number of low and high stimuli) and biased (1:3 and 3:1 ratios of low to high

stimuli) blocks of trials. Before beginning each block, we informed the subjects how many low and high stimuli would be present.

- Condition 2: Payoff structure.

We manipulated the payoff structure to mimic the stimulus frequency manipulation in Condition 1, setting the decision cutoff at 50 as well. Prior to beginning each block, the subjects were informed of the payoff structure in the upcoming block. In the unbiased blocks, they received 2 points for correctly judging the number of asterisks as low or high and had 2 points deducted when answering incorrectly. In the low-biased blocks, the subjects received 3 points or had 1 point deducted for answering “low” correctly or incorrectly, respectively, and received 1 point or had 3 points deducted for answering “high” correctly or incorrectly, respectively. In the high-biased blocks, the reverse point structure applied: +1/-3 for low responses and +3/-1 for high responses. After each trial, the number of points received or deducted was shown to the subjects; a running score (in points) was shown after each block. We instructed the subjects to maximize their score. In each block, half of the trials consisted of low stimuli and half consisted of high stimuli.

- Condition 3: Decision cutoff.

We manipulated decision cutoff to test whether this manipulation could be differentiated from the stimulus frequency manipulation by the model. We set the cutoff between high and low stimuli at 40, 50, or 60 in Experiment 1 and at 45 or 55 in Experiment 2 (across blocks). The number of low and high stimuli displayed in a block was balanced such that half of the trials consisted of low stimuli and half consisted of high stimuli, regardless of decision cutoff.

- Condition 4: Decision cutoff and stimulus frequency.

If Condition 1 could be differentiated from Condition 3, what might happen in a condition in which those two manipulations were combined? To answer this question, we used the same task as in Condition 3, with one trial from each number of asterisks in each block (as in Condition 2), such that the blocks in which the decision cutoff was set at the lowest value contained $\frac{1}{4}$ of low stimuli and $\frac{3}{4}$ of high stimuli and, conversely, the blocks in which the decision cutoff was set at the highest value contained $\frac{3}{4}$ of low stimuli and $\frac{1}{4}$ of high stimuli.

2.4 Modeling analysis

We fit the diffusion model to the response proportions and RT distributions for correct and error responses and obtained parameter values that produce predicted values (for response proportions and RT distributions) that were as close as possible to the data. In both experiments, the diffusion model was fit to data from individual subjects. Extreme fast and slow responses were eliminated from analyses using cutoffs.

In Experiment 1, lower and upper cutoffs were determined by analysis of 50-ms windows to eliminate fast responses at or below chance and slow responses.² Lower cutoffs were 240 ms and 270 ms for Conditions 1 and 2, respectively; upper cutoffs were 1500 ms in Condition 1 and 1700 ms in Condition 2. As a result, approximately 1.4% of the data in Condition 1 and 1.6% of the data in Condition 2 were excluded. In Conditions 3 and 4, the cutoffs were 100 ms and 3000 ms (i.e., approximately 0.7% of the data in Condition 3 and 0.8% of the data in Condition 4).³ RT data were divided into four levels of difficulty in eight groups, determined by the number of asterisks present in each stimulus, as follows: 31–35 or 66–70; 36–40 or 61–65; 41–45 or 56–60; 46–50 or 51–55. RT distributions were approximated by five quantiles, evenly spaced between .1 and .9 (hence representing the times at which 10%, 30%, 50%, 70%, and 90% of the correct or error responses were terminated in each condition). In Experiment 2, we used more conservative cutoffs such that no more than 0.5% of the data points of each subject were excluded. Lower and upper cutoffs were 100 ms and 1500 ms,⁴ respectively (except for Subject C, whose upper cutoff was extended to 1900 ms such that only 0.49% of the data points were excluded).

The method we used to fit the diffusion model was the chi-square method, which works as follows (cf. Ratcliff & Tuerlinckx, 2002). First, theoretical (simulated by the model) and empirical RT data were grouped into six bins, separately for correct and error responses. These bins were defined by five evenly spaced quantiles, namely, .1, .3, .5, .7, and .9, producing two extreme bins each containing 10% of the observations and the others each containing 20%. Inserting the quantile RTs for correct or error responses into the cumulative probability func-

tion gives the expected cumulative probability up to that quantile for the respective responses. Subtracting the cumulative probabilities for each successive quantile from the next higher quantile gives the proportion of responses expected between adjacent quantiles, and multiplying by the total number of observations (correct or error responses) gives the expected frequencies in each bin.⁵ Summing over $(observed - expected)^2 / expected$ for all conditions gives the chi-square statistic (χ^2) to be minimized by model-parameter adjustment. The minimization routine we used was based on the SIMPLEX fitting method (Nelder & Mead, 1965), which takes a set of initial parameter values and adjusts them to minimize the χ^2 (for a more detailed explanation of the SIMPLEX routine, see Ratcliff & Tuerlinckx, Appendix B).

To test our three hypotheses, we planned (pairwise) nested-model comparisons, based on the χ^2 -statistic produced by different model variants in fitting the same data. For example, to test H2, whether stimulus frequency affects starting point (but not drift criterion) and decision cutoff affects drift criterion (but not starting point), we fit a model in which both starting point and drift criterion could vary to the data in the stimulus-frequency manipulation. Then we compared the χ^2 -statistic produced by that model to the χ^2 -statistic produced by a model in which starting point (but not drift criterion) could vary and by a model in which drift criterion (but not starting point) could vary. Note that each of these two variants is nested within the initial model. These comparisons showed which model fit the data in the stimulus-frequency manipulation best; repeating them to the data in the decision cutoff showed which model fit those data best, obtaining support for or against H2. The key to this chi-square difference test comparing pairs of models is that the difference in statistics between the two models is asymptotically distributed as a chi-square variate with degrees of freedom given by the difference in the number of parameters estimated under the two models (e.g., Bentler & Bonett, 1980, p. 593). In the case of H2 in Experiment 1, there were two degrees of freedom in each comparison between the first model described above and each of the two variants. The test implies a null hypothesis of equality in parameter estimates (between models) versus an alternate hypothesis of inequality. If the chi-square variate exceeds the critical value in the test (9.21 at the 99% confidence level for two degrees of freedom),⁶ the null hypotheses is rejected in favor of the alternate.

²Starting at 100 (or 2000) ms, we measured accuracy of responses within the next (or prior) 50 ms. If accuracy within that time window was above chance, then the lower (or upper) cutoff was set to the lower (or upper) limit of that time window. Otherwise, we examined the next time window and repeated that operation until above-chance accuracy was observed.

³In practice, the lower cutoff could have been set at 200 ms, as there were only .03% and .01% of the responses between 100 ms and 200 ms in Conditions 3 and 4, respectively.

⁴As in Experiment 1 (Conditions 3 and 4), in practice, the lower cutoff could have been set at 200 ms, as there were only .08% of the responses between 100 ms and 200 ms.

⁵If there were fewer than five errors in an experimental condition, five quantiles could not be computed, and the error reaction times for the condition were excluded from the chi-square computation.

⁶We adopted the 99% confidence level to disambiguate model parameter influence, rather than the commonly used 95% confidence level for experimental effects, because of the growing concern that null hypothesis significance tests using $p = .05$ may not lead to strongly supported conclusions (e.g., Kruschke, 2011; Wetzel et al., 2011).

2.4.1 Model fits

Data and model predictions are displayed as quantile probability functions throughout this article (e.g., Figure 4, for Condition 1 in Experiment 1). For each difficulty condition in the experiment, the .1, .3, .5 (median), .7, and .9 quantiles of the RT distribution are plotted as a function of response proportion. That is, the quantiles are plotted as five points stacked along the y-axis at a response proportion point on the x-axis. Thus, the shape of the RT distribution is given by the distance between pairs of quantiles (e.g., larger distances between .5 and .9 than between .1 and .5 quantiles represent right skewness, typically present in choice-RT human data). To distinguish data points from model predictions readily, data points are plotted as circles and model predictions are plotted as crosses, connected across difficulty conditions by a dashed line indicating matching quantiles (i.e., the line closest to the x-axis connects all .1 quantiles, whereas the furthest line connects all .9 quantiles).

The proportions on the right of each panel (toward 1) in the quantile-probability plots come from correct responses, and the proportions on the left come (toward 0) from error responses. This is so because whereas an extremely difficult condition is expected to produce correct responses just above 50% of the time, a very easy condition is expected to produce correct responses nearly 100% of the time; proportion of error responses complement those of correct responses. Further, because we present low and high responses separately, correct low responses plus error high responses constitute all responses in the conditions for which low responses should have been made, and vice-versa. We chose quantile probability functions over other ways of displaying the data because they show information about all the data: the proportions of correct and error responses, and the shapes of the RT distributions for correct and error responses. This means, for example, that the change in shape of RT distributions can be examined as a function of difficulty, and comparisons can be made between correct and error responses.

3 Experiment 1

We designed Experiment 1 to examine whether stimulus frequency and payoff structure affect the same component parameters driving the decision in the same way (H1) and whether stimulus frequency affects starting point in the model, whereas decision cutoff affects drift criterion (H2 and H3). We show below that we found that stimulus frequency and payoff structure did not affect the same component parameters driving the decision in the same way, but that manipulations of stimulus frequency affect starting point only and manipulations of decision cutoff affect drift criterion only.

3.1 Method

3.1.1 Subjects

Twelve Ohio State University students were tested in four conditions: four subjects in Condition 1, four subjects in Condition 2, and four subjects in Conditions 3 and 4. Subjects were tested during five (Conditions 1 and 2) or four (Conditions 3 and 4) sessions of approximately 45 to 55 mins each. In Conditions 1, 3, and 4, they were compensated at the rate of \$10.00 per session; in Condition 2, they received \$10.00 per session plus a bonus compensation contingent upon performance, which averaged \$1.55 per session.

3.1.2 Design

Each experimental session was composed of a warmup block followed by 38 blocks (in Conditions 1 and 2—Stimulus frequency and payoff structure, respectively) or 39 blocks (in Conditions 3 and 4—Decision cutoff and decision cutoff with stimulus frequency, respectively) of 40 trials each. Subjects were informed of the block structure prior to the beginning of each block, and blocks were presented in random order for each subject. In Condition 2, we informed the subjects we would reward them one bonus cent for each 15 (positive) points they accumulated throughout the study.

3.2 Results & discussion

As discussed in Section 2.4, we fit three models to the data in each of the four levels of the task manipulation (viz., stimulus frequency, payoff structure, decision cutoff, and decision cutoff along with stimulus frequency). For each condition, there were 264 degrees of freedom in the data ($24 \text{ conditions} \times (12 \text{ bins} - 1)$). The most flexible model, referred to as the full model, had 18 free parameters (two of which varied across bias conditions: drift criterion and starting point), as follows: boundary separation; nondecision time and its range; three starting points (z) at the three bias levels; range in starting point; variability in drift rate; eight mean drift rates at each of the eight groups determined by the number of asterisks in each stimulus; and two drift criteria (d_c), one for each of the two non-neutral bias conditions. The two drift criteria are added to the eight mean drift rate estimates at the neutral condition to produce the mean drift rate estimates at the respective bias condition. The (fixed) scaling parameter, s , was set to 0.1, as is usual (e.g., Ratcliff & Rouder, 2000). The other two models, nested in the full model, each had 16 free parameters, either keeping the same starting point at all three bias levels (the same- z model) or setting d_c at 0 (the no- d_c model).

Below, we present the results of our modeling analysis for each of the task manipulations and provide initial discussions based on them. The summary of our findings is as follows: starting point was the critical model parameter in accounting for changes in stimulus frequency; the payoff structure manipulation did not produce the same results as the stimulus frequency manipulation (i.e., starting point was less affected by payoffs than by stimulus frequency, and a shift in drift rates helped the fit across payoffs but not across stimulus frequencies); shift in drift rates, rather than in starting point, was critical for the model to account for changes in decision cutoff; and shift in drift rates, coupled with shift in starting point, was necessary for the model to account for changes in decision cutoff and stimulus frequency simultaneously.

3.2.1 Condition 1: Stimulus frequency

The data in this condition showed that subjects favored the most frequent alternative over the least frequent one. In Figure 4, this can be observed by comparing the data points in left panels to those in right panels across bias conditions. Whereas there was no clear difference in the neutral condition (top row), low responses were faster and more accurate than high responses when low stimuli appeared more frequently than high stimuli (middle row; see four right-most stacks of five circles in both left and right panels), and high responses were faster and more accurate than low responses when high stimuli appeared more frequently than low stimuli (bottom row). In addition, responses to frequent stimuli were also faster and more accurate than responses to equally likely stimuli (top-left vs. middle-left panels; top-right vs. bottom-right panels).

As shown in Table 1 (Columns d_c^l and d_c^h), parameter estimates in the full model (averaged across subjects), resulted in near-zero drift criterion values, suggesting d_c was not a necessary parameter to account for the behavioral effects due to our manipulation of stimulus frequency. A nested-model comparison on individual parameter estimates between the full model and the no- d_c model showed that the latter was able to fit the data statistically as well as the full model for two of the four subjects. A nested-model comparison between the full model and the same- z model showed that the former was a much better fit to the data for all subjects (Table A1). Thus, changing starting points across conditions of stimulus frequency was the crucial change to account for the effects in the data. The changes in starting-point estimates were as systematic as the changes in RT and accuracy in the data: in comparison to the neutral condition, starting point moved closer to the boundary associated with low responses for the condition in which low stimuli were more frequent than high stimuli and moved away

from it for the condition in which high stimuli were more frequent than low stimuli.

Based on these results, we favor a model in which only starting point is allowed to vary in response to manipulations of stimulus frequency, consistent with our hypothesis (H2). (The predictions in Figure 4 illustrate the no- d_c model fits.) We rejected the model in which starting point is fixed but mean drift rates vary across bias conditions because it was not competitive to fit the data; we rejected the model in which both starting point and mean drift rates vary across bias conditions because it is less parsimonious (than the no- d_c model) and when it significantly improved the fits over the no- d_c model, it did so with fairly small changes in mean drift rates, viz., $|0.023|$, averaged across bias conditions for Subjects 2 and 4.

3.2.2 Condition 2: Payoff structure

Unlike in Condition 1, in contrast with the baseline condition, the payoffs favoring one alternative over the other produced only a slight gain in accuracy and no strong shift in the RT distributions. In Figure 5, responses to stimuli associated with high reward and low cost were not noticeably faster or more accurate than responses to equally likely stimuli (top-left vs. middle-left panels; top-right vs. bottom-right panels).

In the model, averaged parameter estimates resulted in non-zero drift criterion values across payoff conditions and smaller effects of payoff on starting point values (Table 2). This difference in the effects due to payoffs vs. the effects due to stimulus frequency is contrary to our hypothesis (H1). A nested-model comparison on individual parameter estimates between the full model and the no- d_c model showed that the former was a statistically better fit to the data for three of the four subjects. A nested-model comparison between the full model and the same- z model showed that the former was a statistically better fit to the data for two of the four subjects (Table A2).

Thus, unlike in Condition 1, we found no strong support for or against the nested models. That is, the payoff structure manipulation produced behavioral effects that were accounted for with changes in both starting point (z) and quality of evidence (drift criterion, d_c) for data from two subjects, but were accounted for by changes in z only for one subject and by changes in d_c only for another. Specifically, as in Condition 1, when the no- d_c model fit the data statistically as well as the full model (Subject 8), the starting point moved closer to the boundary associated with high responses for trials in which high stimuli were rewarded more highly than low stimuli and away from it for the opposite payoff condition. This change in starting point, however, was small in comparison to the changes in starting point observed in Condition 1 (cf. Table A1). When the same- z model fit the data about as well as the

Figure 4: Quantile-Probability plots for data and no- d_c model predictions from Condition 1. Quantile-RT data points, averaged across subjects, are plotted in ascending order, from .1 to .9 in each column of circles. In each panel, reading it from left to right, there are four such columns across error responses followed by four columns across correct responses, making up the eight difficulty conditions (for asterisk counts of 31–35, 36–40, 41–45, 46–50, 51–55, 56–60, 61–65, and 66–70). The horizontal position at which each quantile-RT column is plotted is the response proportion corresponding to that difficulty condition. Model predictions are plotted as crosses, connected at the same quantile levels across difficulty conditions. Note that error quantiles in conditions with fewer than five observations (for each subject) could not be computed. For error columns with eleven or fewer data points, only the median RT was plotted (excluding subjects with no error responses) to indicate the level of accuracy in those conditions (as a diamond).

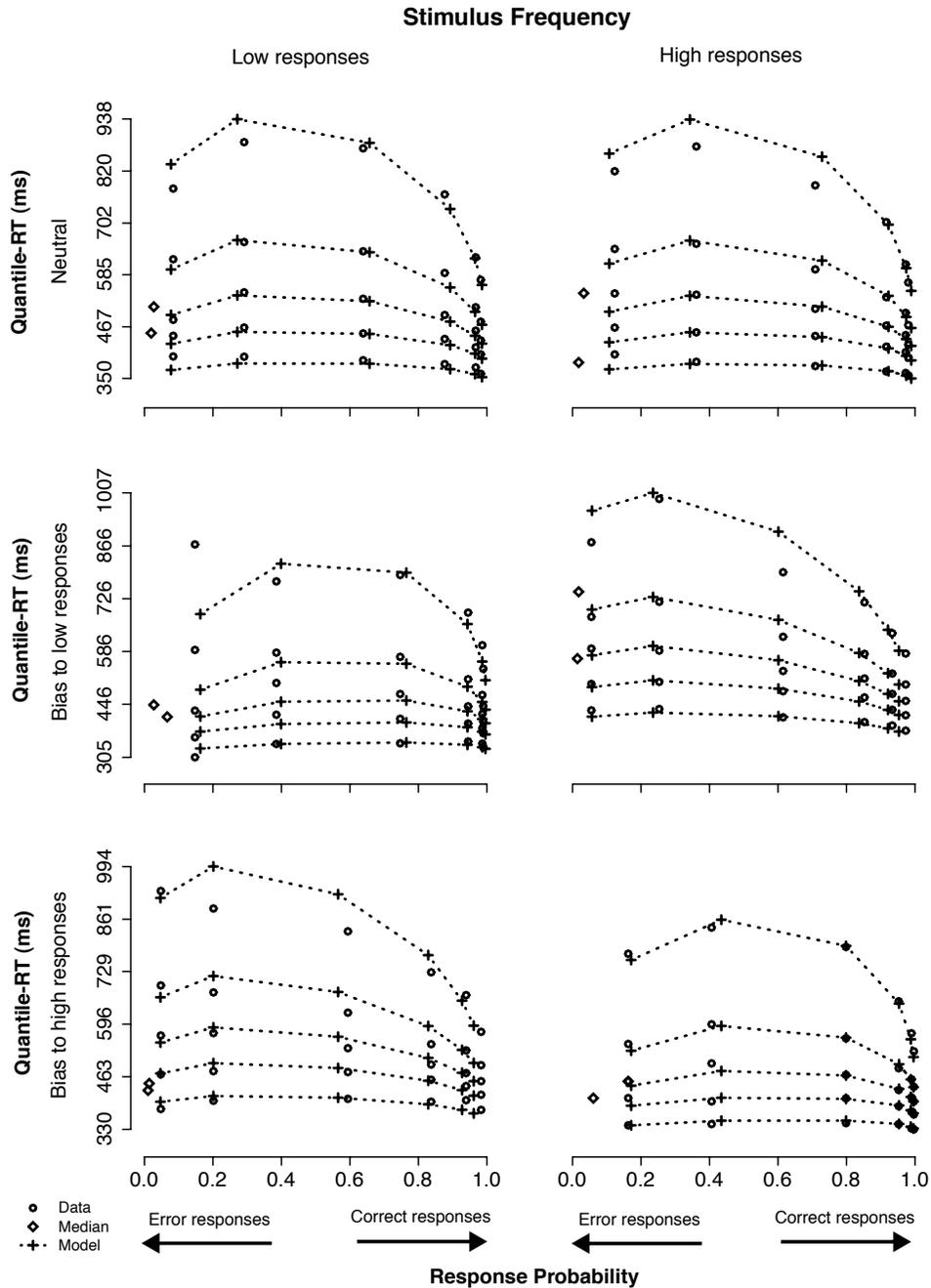


Table 1: Mean parameter estimates: Stimulus frequency, Experiment 1.

Full model (18 free parameters)									
a	T_{er}	s_t	z_l	z_n	z_h	s_z	d_c^l	d_c^h	
0.117	0.329	0.119	0.077	0.058	0.044	0.055	0.004	0.008	
η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
0.142	0.516	0.408	0.270	0.085	-0.137	-0.316	-0.448	-0.543	395

Note. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z = starting point at trials with no bias, or bias toward low or high responses); s_z = range of variability in starting point; d_c = drift criterion at the low or high bias condition; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup.

Table 2: Mean parameter estimates: Payoff structure, Experiment 1.

Full model (18 free parameters)									
a	T_{er}	s_t	z_l	z_n	z_h	s_z	d_c^l	d_c^h	
0.137	0.375	0.118	0.070	0.066	0.063	0.049	0.026	-0.043	
η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
0.155	0.597	0.460	0.283	0.093	-0.109	-0.304	-0.472	-0.583	375

Note. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z = starting point at trials with no bias, or bias toward low or high responses); s_z = range of variability in starting point; d_c = drift criterion at the low or high bias condition; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup.

full model (Subjects 5 and 8), drift criterion values produced shifts in mean drift rates toward high responses for trials in which high stimuli were rewarded more highly than low stimuli, and in the opposite direction for the opposite payoff condition (Table A2).

In summary, we could not favor the model in which starting point is allowed to vary in response to manipulations of payoff structure, as we did for manipulations of stimulus frequency, contrary to our hypothesis (H1). (For comparison with Condition 1, the predictions in Figure 5 illustrate the no- d_c model fits.) Reflecting the little to no change in RT distributions and only slight improvement in accuracy for conditions in which one of the two alternatives was associated with a better payoff than the other, the model was able to account for the data with relatively small changes in starting point and mean drift rates. For two of the four subjects, the model with both starting point and mean drift rates varying across bias conditions produced statistically better fits than either nested model. For another, we observed that the nested models were statistically indistinguishable. As a result, we could not rule out the full model for this condition.

3.2.3 Condition 3: Decision cutoff

The data in this condition showed that subjects adjusted to our manipulation of decision cutoff with little to no

change in RT distributions but with noticeable changes in accuracy associated with stimuli made out of the same number of asterisks, reflecting a difficulty change implied by the change in decision cutoff. In Figure 6, this can be observed by comparing the correct data points in the top row with the correct data points in either the middle or the bottom rows.

In the model, averaged parameter estimates showed neither systematic nor large changes in starting point across decision cutoffs, suggesting starting point was not particularly sensitive to the behavioral effects of our manipulation of decision cutoff. In addition, drift criterion estimates were, on average, four times as large as in Condition 2 and approximately seven times as large as in Condition 1 (Table 3). A nested-model comparison on individual parameter estimates between the full model and the no- d_c model showed that the former was a much better fit to the data for all subjects. A nested-model comparison between the full model and the same- z model showed that the latter was able to fit the data statistically as well as the full model for two of the four subjects (Table A3). Thus, shifts in mean drift rates across decision-cutoff conditions were the crucial change to account for the effects in the data: in comparison to the neutral condition, as decision cutoff was moved higher, drift rates were shifted such that very-low stimuli were identified

Figure 5: Quantile-Probability plots for data and no- d_c model predictions from Condition 2. Quantile-RT data points, averaged across subjects, are plotted in ascending order, from .1 to .9 in each column of circles. In each panel, reading it from left to right, there are four such columns across error responses followed by four columns across correct responses, making up the eight difficulty conditions (for asterisk counts of 31–35, 36–40, 41–45, 46–50, 51–55, 56–60, 61–65, and 66–70). The horizontal position at which each quantile-RT column is plotted is the response proportion corresponding to that difficulty condition. Model predictions are plotted as crosses, connected at the same quantile levels across difficulty conditions. Note that error quantiles in conditions with fewer than five observations (for each subject) could not be computed. For error columns with eleven or fewer data points, only the median RT was plotted (excluding subjects with no error responses) to indicate the level of accuracy in those conditions (as a diamond).

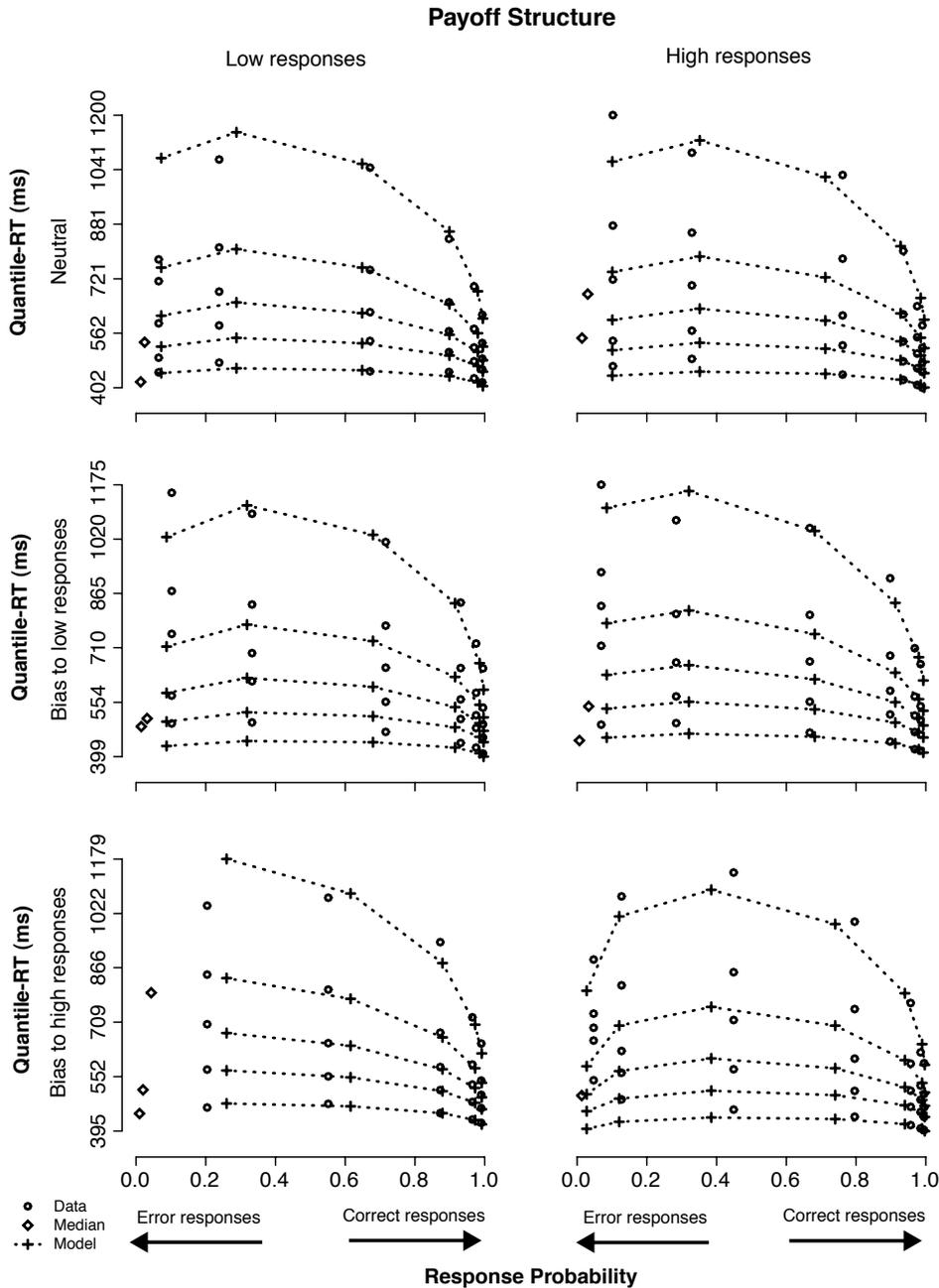
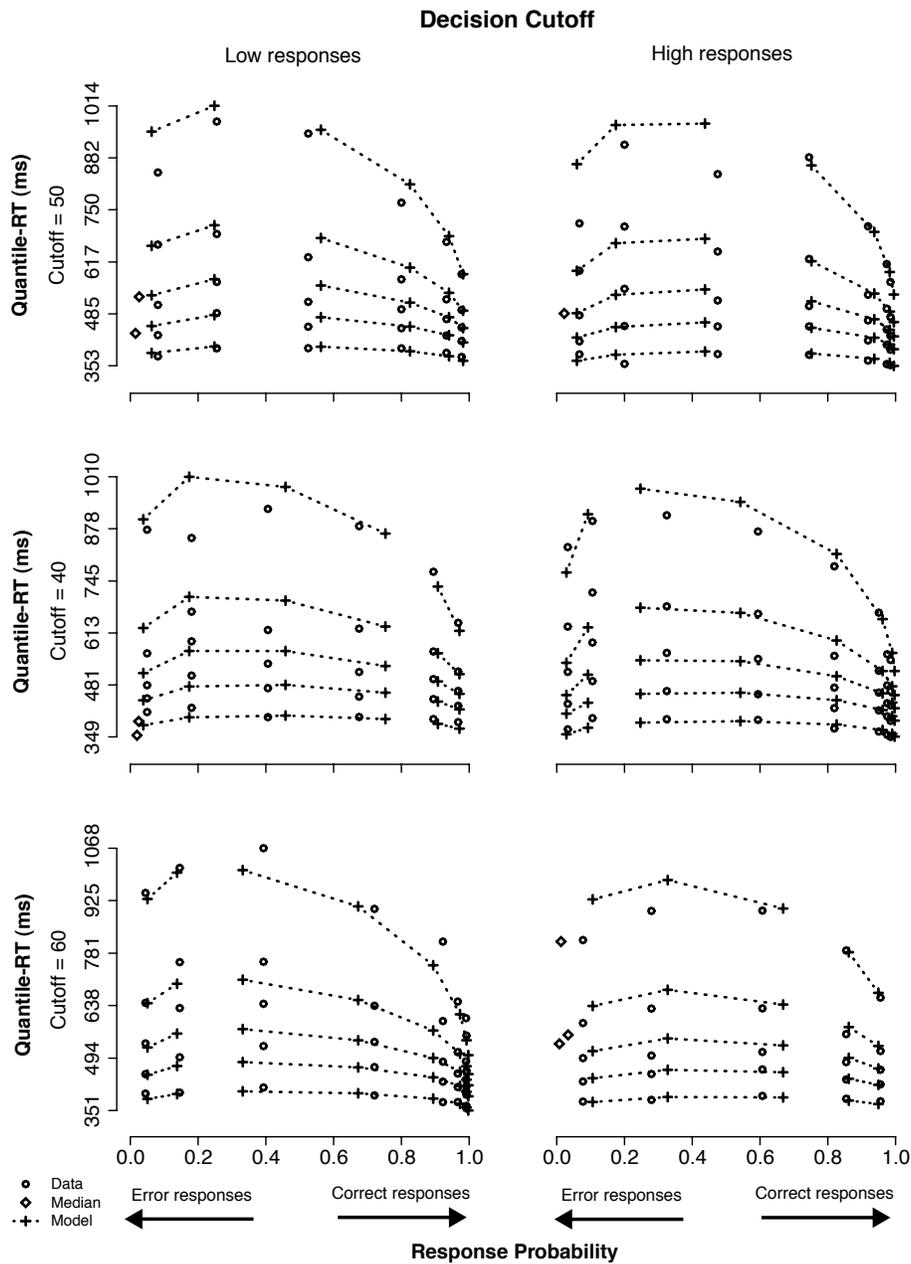


Figure 6: Quantile-Probability plots for data and same- z model predictions from Condition 3. Quantile-RT data points, averaged across subjects, are plotted in ascending order, from .1 to .9 in each column of circles. In each panel, reading it from left to right, there are four such columns across error responses followed by four columns across correct responses, making up the eight difficulty conditions (for asterisk counts of 31–35, 36–40, 41–45, 46–50, 51–55, 56–60, 61–65, and 66–70). The horizontal position at which each quantile-RT column is plotted is the response proportion corresponding to that difficulty condition. Model predictions are plotted as crosses, connected at the same quantile levels across difficulty conditions. Note that error quantiles in conditions with fewer than five observations (for each subject) could not be computed. For error columns with eleven or fewer data points, only the median RT was plotted (excluding subjects with no error responses) to indicate the level of accuracy in those conditions (as a diamond). Discontinuation of dotted lines emphasize the separation between correct and error responses.



more easily; as decision cutoff was moved lower, drift rates were shifted such that very-high stimuli were identified more easily.

Based on these results, we favor a model in which only drift criterion is allowed to vary in response to manipulations of decision cutoff, consistent with our hypothesis (H2). (The predictions in Figure 6 illustrate the same- z model fits.) We rejected the model in which starting point varied but mean drift rates were fixed across conditions because it was not competitive to fit the data; we rejected the model in which both starting point and mean drift rates varied across conditions because it is less parsimonious (than the same- z model) and when it significantly improved the fits over the same- z model, it did so with fairly small changes in starting point estimates between non-neutral conditions, viz., 9% (averaged across Subjects 9 and 10).

3.2.4 Condition 4: Decision cutoff and stimulus frequency

The data in this condition showed that subjects adjusted to our simultaneous manipulation of decision cutoff and stimulus frequency with changes in the RT distributions. For example, in Figure 6, comparing the neutral condition to the condition in which the cutoff is set at 60, approximately a 10% shorter leading edge (about 39 ms) and a 5% longer tail (about 36 ms) were observed in the correct RT distribution for low responses (on the average across the four levels of difficulty past the corresponding cutoff; top-left panel vs. bottom-left panel). This finding is sensible because our experimental manipulation produced blocks with unequal numbers of high and low responses in the non-baseline (experimental) conditions. For example, when the decision cutoff was set at 40, there were 10 low-response trials and 30 high-response trials in a block, as opposed to 20 each in the baseline condition (i.e., decision cutoff set at 50). (Conversely, there were 30 low-response trials and 10 high-response trials in each block in which the criterion was 60.)

The type of distribution change we observed can be accounted for by the model with shifts in either starting point or drift criterion across conditions (Ratcliff, 2002). In the model, averaged parameter estimates showed systematic changes in both starting point drift criterion, suggesting both z and d_c were sensitive to the behavioral effects of our simultaneous manipulation of decision cutoff and stimulus frequency. A nested-model comparison on individual parameter estimates between the full model and both the no- d_c the same- z models showed that the full model was a much better fit to the data for all subjects (Table A4). Thus, shifts in mean drift rates and changes in starting point were both crucial to account for the effects in the data across decision cutoff plus stimulus frequency

conditions.

Based on these results, we favor a model in which both drift criterion and starting point are allowed to vary in response to simultaneous manipulations of decision cutoff and stimulus frequency, consistent with our hypothesis (H3). (The predictions in Figure 7 illustrate the full model fits.) We rejected the model in which starting point varied but mean drift rates were fixed across conditions because it was not competitive to fit the data; likewise, we rejected the model in which starting point was fixed but mean drift rates varied across conditions because it was not competitive to fit the data.

3.2.5 Experiment 1 summary

In Experiment 1, there were four subjects in each condition and a total of twelve subjects across all four conditions. Results showed the following:

- In line with previous findings, starting point was the crucial model parameter to account for changes in stimulus frequency;
- The payoff structure manipulation did not produce the same results as the stimulus frequency manipulation—starting point was less affected by payoffs than by stimulus frequency, and a shift in drift rates helped the fit across payoffs but not across stimulus frequencies;
- Shifts in drift rates, rather than in starting point, were crucial to account for changes in decision cutoff—confirming the interpretation of the “drift criterion” parameter; and
- Shifts in both drift rates and starting point were necessary to account for changes in decision cutoff and stimulus frequency simultaneously.

Experiment 1 results also showed individual differences in performance and model-parameter estimates. Such individual differences are not surprising because there is no reason to believe everyone should perceive a stimulus in the same way or use the same subjective criterion to evaluate stimulus evidence. Nevertheless, we designed Experiment 2 to examine the same manipulations as in Experiment 1 in a within-subjects design.⁷

4 Experiment 2: Within-subjects replication

Our goal in Experiment 2 was to reexamine the interpretation of changes in model-parameter estimates due

⁷We thank Jon Baron for suggesting this replication.

Table 3: Mean parameter estimates: Decision cutoff, Experiment 1.

Full model (18 free parameters)									
a	T_{er}	s_t	z_{40}	z_{50}	z_{60}	s_z	d_c^{40}	d_c^{60}	
0.129	0.344	0.121	0.062	0.058	0.064	0.074	-0.093	0.260	
η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
0.231	0.624	0.454	0.264	0.076	-0.152	-0.378	-0.552	-0.705	508

Note. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z = starting point at decision cutoff equal to 40, 50, or 60; s_z = range of variability in starting point; d_c = drift criterion at decision cutoff equal to 40, 50, or 60; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup.

Table 4: Mean parameter estimates: Decision cutoff and stimulus frequency, Experiment 1.

Full model (18 free parameters)									
a	T_{er}	s_t	z_{40}	z_{50}	z_{60}	s_z	d_c^{40}	d_c^{60}	
0.120	0.319	0.124	0.045	0.056	0.072	0.049	-0.140	0.180	
η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
0.104	0.449	0.355	0.216	0.065	-0.099	-0.237	-0.363	-0.487	669

Note. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z = starting point at decision cutoff equal to 40, 50, or 60; s_z = range of variability in starting point; d_c = drift criterion at decision cutoff equal to 40, 50, or 60; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup.

to manipulations of stimulus frequency, payoff structure, and decision cutoff in a within-subjects design. To produce enough data from all conditions in relatively few sessions, we reduced the number of trials in each condition. In comparison to Experiment 1, we eliminated the neutral condition and kept two bias levels. In addition, we eliminated the easiest trials to reduce the number of difficulty conditions from four to three. As a result, we ran Experiment 2 in five sessions per subject, the first of which was used for subjects to familiarize themselves with the experiment and was not analyzed, and in each of which all conditions were presented.

The modeling analysis was done in a similar manner to that of Experiment 1, noting that each model variant was fit to data from all conditions simultaneously, in response to the within-subjects design. As we show below, one disadvantage of that adjustment was that not all tested models could be designed as nested variants of one another. Nevertheless, in the end of this section we show that we found evidence to support the following Experiment 1 findings: changes in starting point in the model were crucial to account for changes in stimulus frequency across blocks; in comparison to these changes due to stimulus frequency, smaller changes in starting point were needed to account for changes in payoffs; shifts in drift rates in the model, rather than changes in starting point, were

crucial to account for changes in decision cutoff across blocks; and changes in decision cutoff and stimulus frequency simultaneously across blocks were accounted for in the model by simultaneous shifts in drift rates and in starting point.

4.1 Method

4.1.1 Subjects

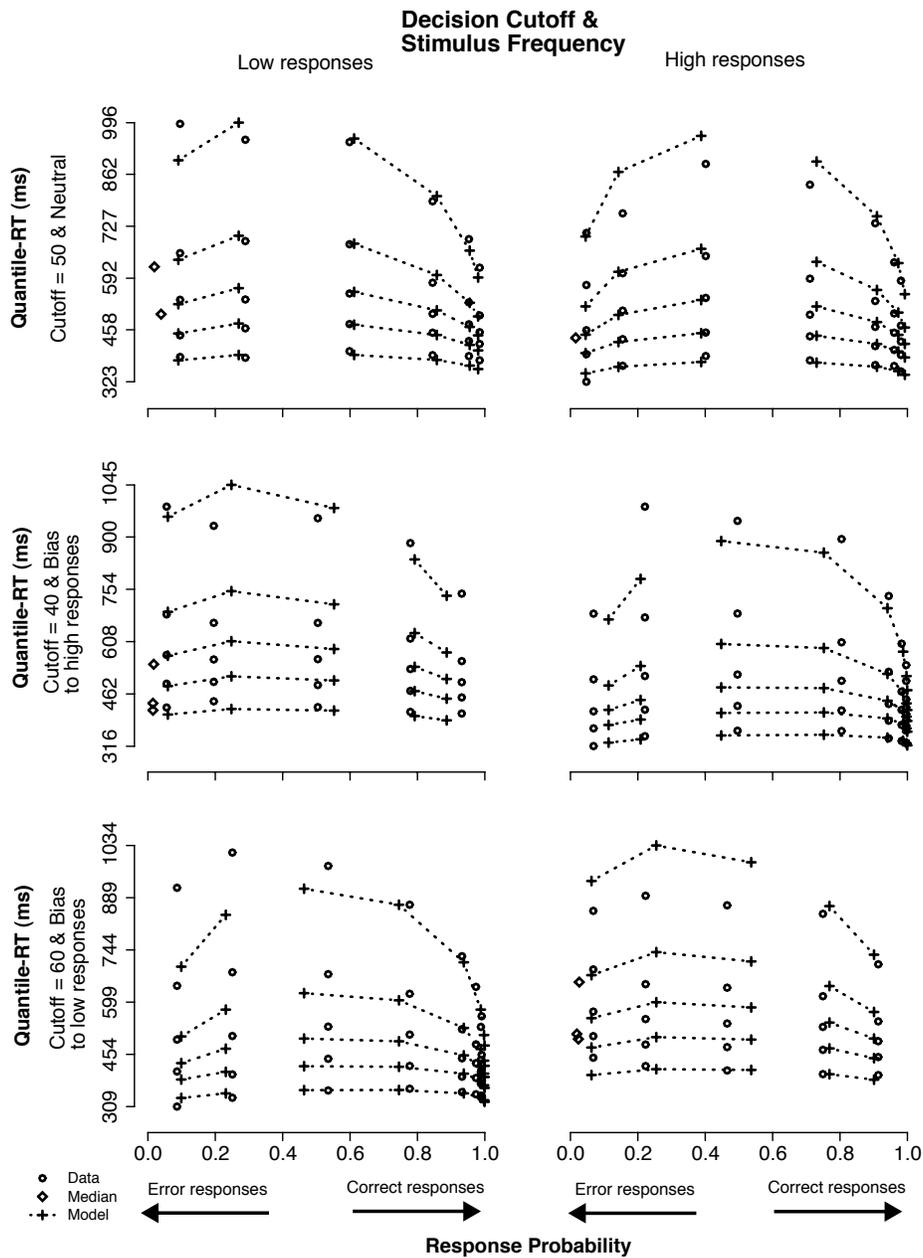
Six Ohio State University students completed five experimental sessions. Each session lasted approximately 40 to 45 minutes, and subjects were compensated at the rate of \$8.00 per session plus a bonus compensation contingent upon performance in the payoff structure condition (for performance between sessions 2 and 5; average = \$1.09 per session).

4.1.2 Design

Experiment 2 was a $4 \times 3 \times 2 \times 2$ within-subjects design, with task, difficulty, bias level, and type of response (low or high) as factors, respectively. In the payoff task, we rewarded the subjects one bonus cent for each 50 (positive) points they accumulated.⁸

⁸The increase from 15 to 50 points, relative to Experiment 1, was due to the threefold increase in payoff and the absence of neutral trials.

Figure 7: Quantile-Probability plots for data and full model predictions from Condition 4. Quantile-RT data points, averaged across subjects, are plotted in ascending order, from .1 to .9 in each column of circles. In each panel, reading it from left to right, there are four such columns across error responses followed by four columns across correct responses, making up the eight difficulty conditions (for asterisk counts of 31–35, 36–40, 41–45, 46–50, 51–55, 56–60, 61–65, and 66–70). The horizontal position at which each quantile-RT column is plotted is the response proportion corresponding to that difficulty condition. Model predictions are plotted as crosses, connected at the same quantile levels across difficulty conditions. Note that error quantiles in conditions with fewer than five observations (for each subject) could not be computed. For error columns with eleven or fewer data points, only the median RT was plotted (excluding subjects with no error responses) to indicate the level of accuracy in those conditions (as a diamond). Discontinuation of dotted lines emphasize the separation between correct and error responses.



Each of the six subjects who completed all five sessions completed no more than one session a day. Time to completion of all sessions ranged from six to ten days. From session to session, we alternated the response-key mapping (i.e., “Z” was associated with low-number responses in one session and with high-number responses in the following session) to minimize the potential of subjects speeding up responses from session to session.⁹ All 48 conditions were present in each session: task and bias level were manipulated across blocks, and difficulty and response were manipulated within blocks. Each experimental session was composed of 64 blocks (16 for each task) of 36 trials each. We informed the subjects of the block structure prior to the beginning of each block, and we used a Latin-square design to counterbalance the order of presentation of the four tasks across sessions; in a session, all blocks with a particular task were presented sequentially. We summarize the levels of each of the four task conditions below, for each of which the number of asterisks in each stimulus ranged from 36 to 65.

- Condition 1: Stimulus frequency. With the decision cutoff set at 50, we manipulated the proportion of low to high stimuli such that there were (biased) blocks with 1:3 and 3:1 ratios of low to high stimuli.
- Condition 2: Payoff structure. We manipulated the payoff structure to mimic the stimulus frequency manipulation in Condition 1. In the low-biased blocks, the subjects received 27 points or had 9 points deducted for answering “low” correctly or incorrectly, respectively, and received 9 points or had 27 points deducted for answering “high” correctly or incorrectly, respectively. In the high-biased blocks, the reverse point structure applied: +9/-27 for low responses and +27/-9 for high responses. We instructed the subjects to maximize their score and informed them we would reward them one bonus cent for each 50 (positive) points they accumulated. In each block, half of the trials consisted of low stimuli and half consisted of high stimuli.
- Condition 3: Decision cutoff. We set the cutoff between high and low stimuli at 45 or 55 in different blocks of trials. The number of low and high stimuli displayed in a block was balanced such that half of the trials consisted of low stimuli and half consisted of high stimuli, regardless of decision cutoff.
- Condition 4: Decision cutoff and stimulus frequency. We set the decision cutoff at 45 or 55 across

⁹After running several hours of pilot tests and debriefing subjects, the first author learned that task familiarity may lead subjects to believe, correctly or not, they can finish each subsequent session faster than the previous session without compromising performance. Changing the response-key mapping seemed to interfere with that belief.

blocks, as in Condition 2, but with blocks in which the decision cutoff was set at 45 containing 9 low stimuli and 27 high stimuli and, conversely, with blocks in which the decision cutoff was set at 55 containing 27 low stimuli and 9 high stimuli.

4.1.3 Modeling analysis

Experiment 2 differed from Experiment 1 by having all subjects go through all tasks (in each session). Thus, modeling of the data in Experiment 2 must involve model structures with the flexibility to vary drift criterion and starting point across tasks, as well as reasonable assumptions about which processes each individual subject was likely to keep constant. We tested six model structures, each with a different set of assumptions regarding the parameters of the diffusion model that vary across tasks. (Although each of these model structures is an implementation of the diffusion model, for ease of presentation, in what follows, we refer to them as Models I through VI.)

In the models we tested, we assumed subjects required the same amount of evidence to be considered before making a decision throughout the experiment (i.e., across tasks and sessions). (It is also plausible to assume subjects could change that requirement from session to session, but we collapsed the data across sessions assuming any between-session variation would be small in comparison to experimental manipulations.) We also assumed that the motor response does not change across tasks or sessions (i.e., time to depress “Z” or “/”), and that the mean time associated with encoding the stimuli remains constant. Moreover, the evidence available in stimuli of same difficulty for stimulus frequency, payoff structure, or decision cutoff is physically the same, so we assumed mean drift rates in response to stimuli of same difficulty remained unchanged across tasks. As a result, we expected the following parameters to be constant across tasks: a , a measure of how much evidence is considered for a decision to occur, also referred to as a subject’s degree of conservativeness; T_{er} , a measure of the time needed to encode and respond to the stimulus; mean drift rates ($\nu_{1..6}$), mapped to groups of stimuli with similar number of asterisks; and across-trial variability in T_{er} (s_t) and in ν (η).

The six models we fit to the data in Experiment 2 are summarized in Table 5, informing the following:

- Model I assumed that changes in stimulus frequency led to shifts in starting point of the accumulation of evidence toward the decision threshold and that changes in decision cutoff led to shifts of the quality of the evidence extracted from similar physical stimuli. It also assumed that changes in payoff structure led to shifts in starting point identical to the shifts due to stimulus frequency. In addition, it assumed

Table 5: Structure of tested models across task structures.

Model	Fixed across tasks					Stimulus frequency				Payoffs			
	a	T_{er}	s_t	η	$\nu_{1..6}$	z_l^{freq}	z_h^{freq}	s_z^{freq}	d_c^{freq}	z_l^{pay}	z_h^{pay}	s_z^{pay}	d_c^{pay}
I	free	free	free	free	free	free	free	free	0	free	free	free	0
II	free	free	free	free	free	free	free	free	0	free	free	free	0
III	free	free	free	free	free	free	free	free	0	free	free	free	0
IV	free	free	free	free	free	free	free	free	0	free	free	free	0
V	free	free	free	free	free	free	free	free	0	free	free	free	0
VI	free	free	free	free	free	free	free	free	0	free	free	free	0

Model	Decision cutoff					Stimulus frequency + decision cutoff						# free
	z_{55}^{crit}	z_{45}^{crit}	s_z^{crit}	$d_c^{55;crit}$	$d_c^{45;crit}$	z_{55}^{cfr}	z_{45}^{cfr}	s_z^{cfr}	$d_c^{55;cfr}$	$d_c^{45;cfr}$		
I	free	free	free	free	free	free	free	free	free	free	20	
II	free	free	free	free	free	free	free	free	free	free	23	
III	free	free	free	0	0	free	free	free	free	free	24	
IV	free	free	free	free	free	free	free	free	0	0	23	
V	free	free	free	free	free	free	free	free	free	free	22	
VI	free	free	free	free	free	free	free	free	free	free	23	

Note. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; ν_n = drift rate for the n stimulus subgroup; η = variability in ν ; z = starting point under low or high bias conditions or decision cutoff set at 45 or 55; s_z = range of variability in z ; d_c = drift criterion; # free = total of free parameters. Entries indicate whether model parameters are estimated (free), set to 0, or set to be equal to other parameters (matching colors within a row).

that simultaneous changes in stimulus frequency and decision cutoff led to shifts both in starting point and of the quality of the evidence extracted from similar physical stimuli—the latter of which were identical to shifts due to decision cutoff only.

- Model II differed from Model I by assuming that the shifts in starting point due to changes in payoff structure were different from the shifts in starting point due to changes in stimulus frequency.
- Model III differed from Model II by assuming that changes in decision cutoff led to shifts in starting point but not of the quality of the evidence extracted from similar physical stimuli, while simultaneous changes in stimulus frequency and decision cutoff led to shifts both in starting point and of the quality of the evidence.
- Model IV differed from Model II by assuming that simultaneous changes in stimulus frequency and decision cutoff did not lead to shifts of the quality of the evidence extracted from similar physical stimuli.
- Model V differed from Model II by assuming that simultaneous changes in stimulus frequency and decision cutoff did not lead to shifts in starting point of the accumulation of evidence.

- Model VI differed from Model II by assuming that changes in stimulus frequency and in payoff structure led to shifts of the quality of the evidence extracted from similar physical stimuli but not in starting point of the accumulation of evidence.

4.2 Results & discussion

From Experiment 1, we learned that stimulus frequency affected starting point in the model more strongly than did payoff structure and that, whereas stimulus frequency affects starting point, decision cutoff affects drift criterion. Below we show that we replicated these findings in Experiment 2. Specifically, the model that fit the data best assumed that changes in stimulus frequency and in payoff structure produced changes in starting point parameters, that changes in starting point due to stimulus frequency were greater than changes in starting point due to payoffs, that decision cutoff produced shifts in drift rates corresponding to similar physical stimuli, and that simultaneous changes in stimulus frequency and decision cutoff produced changes in both starting point and drift rate parameters. Illustrations of the model fits are shown in Figures A1 through A4, in which we show quantile-RT and accuracy data along with model predictions for stimulus frequency, payoff, decision cutoff, and decision

cutoff combined with stimulus frequency, respectively.

For each individual subject, there were 528 degrees of freedom in the data ($48 \text{ conditions} \times (12 \text{ bins} - 1)$). We fit the six models summarized in Table 5 to the data in Experiment 2, as follows.

We used Models I and II to test whether stimulus frequency and payoff structure manipulations produced similar changes in the component parameters driving the decision (H1). Both models assumed that changes in decision cutoff led to shifts of the quality of the evidence present in similar physical stimuli. Model II assumed that payoff structure and stimulus frequency manipulations produced different response-bias effects, governed by z . Model I, nested in Model II, assumed that payoff structure and stimulus frequency manipulations produced identical response-bias effects, governed by identical z estimates across the two tasks.

We used Models III and VI to test whether stimulus frequency affects only starting point and decision cutoff affects only drift criterion (H2). Model III assumed that changing decision cutoff also led to changes in z . Model VI assumed that payoff structure and stimulus frequency manipulations produced effects driven by shifts of the quality of the evidence present in similar physical stimuli (governed by parameter d_c , rather than by z).

We used Models IV and V to test whether simultaneous manipulation of stimulus frequency and decision cutoff affected both starting point and drift criterion (H3). Model IV assumed that changing decision cutoff and stimulus frequency simultaneously did not lead to noticeable shifts of the quality of the evidence present in similar physical stimuli because such manipulation would make the stimulus frequency effect stand out. On the other hand, Model V assumed that changing decision cutoff and stimulus frequency simultaneously did not lead to changes in z because such manipulation would make the decision cutoff effect stand out.

As shown in Table 6, note that Models I and V are nested in Model II. Models III, IV, and VI are at least as flexible as Model II, but they are not nested in Model II; we eliminated these models in favor of Model II based on the larger χ^2 -values they produced.

4.2.1 Testing H1

We compared Models I and II to test the assumption that the effects of payoff structure and stimulus frequency manipulations could be modeled by the same parameters. As shown in Table 6 (row 2 vs. row 1), Model II improved fits of Model I by more than 11.3 χ^2 units (the critical value at the 99% level for the nested-model comparison) for all subjects (see Tables A5 and A6 for individual parameter estimates for Models I and II, respectively). As a result, we concluded that the effects of payoff structure

on subjects was different than that of stimulus frequency. Specifically, as shown in Table 7, the average estimates of starting points produced a range in the payoff condition that was 1/3 of the range in the stimulus frequency condition (columns z_l^{freq} vs. z_l^{pay} and z_h^{freq} vs. z_h^{pay}).

4.2.2 Testing H2

We compared Models II and III to test whether the effects of manipulating decision cutoff could be modeled by changes in starting point rather than by shifts in drift rates (Table 6, row 2 vs. row 3). Notice Model III is conceptually different from Model II. In other words, because in Experiment 2 we model across-task data simultaneously, Model III could not be formulated by setting Model II parameters constant or equaling some of Model II parameters. Hence, a nested-model comparison was not possible. Nevertheless, because Model III is more flexible than Model II, Model III would be expected to produce smaller χ^2 values than Model II if both models were true to the data. Rather, we found that our manipulations involving decision cutoff were better fit by shifts in drift rates (by the d_c parameter) as formulated in Model II than by shifts in starting point as formulated in Model III for all subjects (see Tables A6 and A7 for individual parameter estimates for Models II and III, respectively).

Analogously, we comparing Models II and VI (Table 6, row 2 vs. row 6) to test whether the effects of manipulating stimulus frequency and payoff structures could be modeled by shifts in drift rates, rather than changes in starting point. We found that our manipulations of stimulus frequency and comparable payoff effects were best modeled by changes in starting point, except for one subject (see Tables A6 and A10 for individual parameter estimates for Models II and VI, respectively). Taken together, these comparisons showed that the manipulations of stimulus frequency and decision cutoff produced behavioral effects that were accounted for by specific parameters in the model. As shown in Table 7, frequent presentation of low stimuli produced lower estimates of starting point than frequent presentation of high stimuli for a specific set of quality of evidence parameters (columns z_l^{freq} , z_h^{freq} , and ν_{1-6}). On the other hand, changing decision cutoff from high to low produced the same estimate of starting point (approximately half way between estimates in frequent-low and frequent-high conditions) while strengthening the quality of evidence for low and high stimuli, respectively (Table 7, columns z_{55}^{crit} , z_{45}^{crit} , $d_c^{55;crit}$, and $d_c^{45;crit}$).

4.2.3 Testing H3

Having found evidence that changes in stimulus frequency are accounted for by changes in starting point of

Table 6: Model comparison summary: Experiment 2.

Model	# free	Nested	A	B	C	D	E	F	Mean
I	20	Yes	1034.8	1955.3	2025.9	1559.9	934.8	1404.8	1485.9
II	23	-	892.6	1055.5	1602.5	1474.0	906.1	1210.5	1190.2
III	24	No	900.1	1095.8	1696.4	1481.8	968.8	1260.2	1233.9
IV	23	No	905.4	1070.8	1685.0	1496.9	946.7	1232.0	1222.8
V	22	Yes	1305.6	1231.6	1610.9	1496.4	925.5	1229.4	1299.9
VI	23	No	1461.4	2109.6	1721.6	1570.4	905.4	1359.8	1521.4

Note. χ^2 -statistic for each model (in each row) for each subject (across columns A though F). Model number refers to the model structures in Table 5. The number of free parameters in each model is shown on the second column. Comparing Model I to Model II tested the assumption that the effects of payoff structure and stimulus frequency manipulations could be modeled by the same parameters. Model II improved fits of Model I by more than 11.3 for all subjects. Thus, Model II was statistically superior to Model I at the 99% level for all subjects. Comparing Model III to Model II tested whether the effects of manipulating decision cutoff could be modeled by changes in starting point rather than by shifts in drift rates. Because Model III has more free parameters and yet produces poorer (larger) measures of goodness of fit than Model II, we rejected Model III in favor of Model II. Comparing Model VI to Model II tested whether the effects of manipulating stimulus frequency and payoff structures could be modeled by shifts in drift rates, rather than changes in starting point. Because Models II and VI have the same number of free parameters, we interpreted the much poorer (larger) measures of goodness of fit in Model VI for five out of six subjects as evidence that typically stimulus frequency and comparable payoff effects are best modeled by changes in starting point. Comparing Model IV to Model II tested whether the effects of manipulating decision cutoff and stimulus frequency simultaneously could be modeled by changes in starting point only, rather than by changes in starting point and shifts in drift rates. We favored Model II over Model IV because of the better (smaller) measures of goodness of fit it yielded. Comparing Model V to Model II tested whether the effects of manipulating decision cutoff and stimulus frequency simultaneously could be modeled by shifts in drift rates only, rather than by changes in starting point and shifts in drift rates. Model II improved fits of Model V by more than 6.6 for all subjects. Thus, Model II was statistically superior to Model V at the 99% level for all subjects.

Table 7: Mean parameter estimates: Model II.

a	T_{er}	s_t	z_l^{freq}	z_h^{freq}	z_l^{pay}	z_h^{pay}	z_{55}^{crit}	z_{45}^{crit}	
0.108	0.359	0.191	0.035	0.070	0.051	0.062	0.052	0.052	
z_{55}^{cfr}	z_{45}^{cfr}	s_z^{freq}	s_z^{pay}	s_z^{crit}	s_z^{cfr}	d_c^{freq}	d_c^{pay}	$d_c^{55;crit}$	$d_c^{45;crit}$
0.045	0.062	0.050	0.041	0.032	0.031	0	0	-0.047	0.072
$d_c^{55;cfr}$	$d_c^{45;cfr}$	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	χ^2
-0.047	0.072	0.164	-0.359	-0.244	-0.095	0.068	0.221	0.334	1190.2

Note. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z = starting point at low or high bias condition or with decision cutoff set at 45 or 55 for manipulations of stimulus frequency, payoffs, decision cutoff (crit), or decision cutoff and stimulus frequency (cfr); s_z = range of variability in starting point for the corresponding task; d_c = drift criterion; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup. Matching colors indicate that the same parameter was used for both conditions.

the accumulation process but not by changes in the quality of evidence extracted from stimuli and that changes in decision cutoff are accounted for by changes in the quality of evidence extracted from stimuli but not by changes in starting point of the accumulation process, we examined what happened when stimulus frequency and decision cutoff were manipulated simultaneously. Specifically, Model II assumed that this simultaneous manipulation produced changes in both starting points and quality of evidence. Model IV assumed that such manipulation produced no noticeable shifts in the quality of the evidence because it would make the stimulus frequency effect stand out; on the other hand, Model V assumed that it produced no changes in starting point because the decision cutoff effect would stand out.

As shown in Table 6 (row 2 vs. row 4), Model II was superior to Model IV for all subjects (see Tables A6 and A8 for individual parameter estimates for Models II and IV, respectively) in accounting for a simultaneous manipulation of decision cutoff and stimulus frequency. In comparison to Model V (Table 6, row 2 vs. row 5), Model II improved fits to data by more than $6.6 \chi^2$ units (the critical value at the 99% level for the nested-model comparison) for all subjects (see Tables A6 and A9 for individual parameter estimates for Models II and V, respectively). As a result, we concluded that both starting point and drift criterion were sensitive to the simultaneous manipulation of stimulus frequency and decision cutoff. Specifically, as shown in Table 7, changing stimulus frequency and decision cutoff simultaneously produced changes in starting point that were both consistent with and smaller than the changes in starting point due to changes in stimulus frequency only (columns z_l^{freq} and z_h^{freq} vs. columns z_{55}^{cfr} and z_{45}^{cfr})—similar to the changes observed in the payoff condition (cf. columns z_l^{pay} and z_h^{pay})—, while keeping changes in drift criterion equal to the changes observed in the decision cutoff condition.

4.2.4 Experiment 2 summary

Our goal in Experiment 2 was to test the interpretation of changes in model parameters due to changes in stimulus frequency, payoffs, and decision cutoff that we presented in Experiment 1. Results showed that the best-fitting model to the data from individual subjects in Experiment 2 assumed that: changes in stimulus frequency and in payoff structure affected starting point of the accumulation of evidence in the model (i.e., subjects required less evidence for the more frequent or more valued alternative than for the less frequent or less valued alternative); changes in starting point due to stimulus frequency were greater than changes in starting point due to payoffs; decision cutoff produced shifts of the quality of the evidence extracted from similar physical stimuli (modeled

by drift criterion); and simultaneous changes in stimulus frequency and decision cutoff produced changes in both starting point and drift criterion parameters. The parameter estimates of this model are shown in Table A6.

To test the validity of the interpretations derived from the parameter estimates, we performed paired t -tests on the values in Table A6 (across subjects), as follows.

- Starting points in the bias-toward-low condition were significantly lower than starting points in the bias-toward-high condition for stimulus frequency manipulation ($t(5) = -4.146, p = .005$), whereas starting points in the bias-toward-low condition were not significantly lower than starting points in the bias-toward-high condition for payoff structure manipulation ($t(5) = -1.490, p = .098$). As we discussed earlier, this difference was produced by the narrower range of differences between starting points for the two bias conditions: whereas the starting points in the bias-toward-high condition were statistically similar— $t(5) = 1.579, p = .175$, the values in the bias-toward-low condition were significantly distinct— $t(5) = -4.369, p = .007$.
- Starting points in the decision cutoff manipulation fell between bias-toward-low and bias-toward-high values in the stimulus frequency manipulation: they were significantly greater than the bias-toward-low estimates and significantly lower than the bias-toward-high estimates— $t(5) = 4.981, p = .002$; and $t(5) = -3.265, p = .011$, respectively. Drift-criterion estimates for a high decision cutoff (i.e., set at 55) were significantly lower than both estimates for a low decision cutoff (i.e., set at 45) and for an intermediary decision cutoff (i.e., set at 50)— $t(5) = -7.157, p = .000$; and $t(5) = -4.984, p = .002$; estimates for a low decision cutoff were significantly greater than estimates for an intermediary decision cutoff— $t(5) = 6.243, p = .001$.
- In spite of assuming that drift criteria in this simultaneous manipulation would mimic those in the decision cutoff manipulation, starting points for a high decision cutoff were significantly lower than starting points for a low decision cutoff— $t(5) = -3.100, p = .013$. In addition, starting-point estimates in the high decision cutoff condition were significantly lower than estimates in the decision cutoff manipulation, whereas estimates in the low decision cutoff condition were significantly greater than estimates in the decision cutoff manipulation— $t(5) = -2.941, p = .032$; $t(5) = 2.643, p = .046$, respectively.

5 General discussion

In this study, we examined the following four questions: Do manipulations of stimulus frequency and payoff structure affect the same cognitive processes driving the decision and in the same way? How does the manipulation of stimulus frequency compare to that of decision cutoff (threshold)? What happens when stimulus frequency and decision cutoff are manipulated at the same time? Our hypotheses were that: stimulus frequency and payoff structure affect the same component parameters driving the decision in the same way (H1); stimulus frequency affects starting point, whereas decision cutoff affects drift criterion (H2); and, when stimulus frequency and decision cutoff are manipulated at the same time, both starting point and drift criterion are affected (H3).

We reported two experiments involving a two-choice numerosity discrimination in which 18 subjects responded “low” or “high” to the number of asterisks in a 10×10 grid, based on an experimenter-determined decision cutoff. The main manipulation was of task. In the first two conditions, we kept the decision cutoff fixed at 50 and changed the task from stimulus frequency (Condition 1) to payoff structure (Condition 2). In the other two conditions, we manipulated the decision cutoff, either balancing the number of low and high stimulus (Condition 3) or not (Condition 4).

Payoffs and prior probabilities are known to bias choices in near-threshold discrimination tasks. We reported a comparison of how the response bias created by these two manipulations affects specific parameters of the diffusion model. In both experiments, we found that changes in starting point alone were sufficient to account for changes in the RT and accuracy data due to changes in stimulus frequency, consistent with H2 and with previous findings (e.g., Ratcliff et al., 1999). To account for behavioral changes due to changes in payoffs, both starting point and drift criterion changed in Experiment 1, albeit the latter did not do so systematically or strongly; in Experiment 2, we found that changes in starting point alone were sufficient. In both experiments, changes in stimulus frequency produced greater changes in starting point than did changes in payoffs—even in light of payoff changes that mimicked frequency changes. Taken together, these findings provided partial support for H1: stimulus frequency and payoffs affected the same component parameters driving the decision, but the effect was different.

The larger effect of stimulus frequency than of payoff structure on the starting point of the accumulation process assumed to govern the decision is a novel and important empirical result from our study.¹⁰ This suggests that the

two manipulations should not be used interchangeably. The difference between stimulus frequency and payoff structure could be the result of an interference on the processing of the stimuli caused by the presence of the payoff structure (e.g., Diederich & Busemeyer, 2006). Another explanation is that the optimal starting point for biased rewards (i.e., the starting point that would lead to most points) is less biased toward the biased-reward boundary than its analogous stimulus-frequency boundary (contrary to what Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006, p. 734, showed, i.e., it should depend on the fraction of reward for each correct response to one alternative, analogous to changes due to stimulus frequency).

An optimality analysis in the payoff manipulation would have to weight monetary reward against time (or reward rate, the number of correct or rewarded responses per unit of time). Our data suggest that this weighting reduces the effect that might be expected from the payoff manipulation. Generally, a decision about what should be considered optimal would need to be made. For example, in the payoff condition, would maximization of points or maximization of points with RTs below a certain limit be the optimal strategy? Second, response strategies should be examined for each stimulus-difficulty level (governed by the number of asterisks), and our analyses considered all levels simultaneously. Third, in addition to starting point, both boundary separation and drift criterion would also need to be examined to decide on the optimal parameter adjustments. Determining the weighting of two or three variables in an optimality analysis along with time may be a very difficult computational problem (see Starns & Ratcliff, 2010, for an investigation of boundary optimality in the diffusion model).

In the present study, we manipulated the decision cutoff, fixing it at 40, 50, or 60 (Experiment 1), or at 45 or 55 (Experiment 2) across blocks (Conditions 3 and 4). In Condition 3—in which we balanced the number of high and low stimuli—, both experiments showed that changing cutoffs was best fit by shifts in drift rates (governed by the drift criterion parameter, d_c) than by shifts in starting point, consistent with H2. In Condition 4, in which decision cutoff changes also resulted in changes in the number of low and high stimuli, both experiments showed that both starting point and drift criterion parameters of the diffusion model changed with changes in the decision cutoff, consistent with H3. Taken together, these findings show that the empirical signature of a decision cutoff manipulation depends on whether stimulus frequency systematically varies with changes in the decision cutoff. When it does, it should produce shorter leading edges of the RT distributions in the biased condition than in the control condition. Consistent with previous reports (Rat-

¹⁰We know of no prior study that systematically compared a stimulus-frequency manipulation against a similar payoff manipulation, but see Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2011,

for similar smaller effects of payoffs versus stimulus frequency in a fMRI study.

cliff, 1985; Ratcliff et al., 1999), we found that the diffusion model can account for faster leading-edge RTs by moving the starting point of its accumulation process and for constant leading-edge RTs across different decision cutoffs by shifts in its mean accumulation rate.

The overall results provide a systematic account of three experimentally induced response biases in perceptual decision making. They are captured in the framework of the diffusion model, a decision model that accounts for accuracy and RT distributions for correct and error responses. To summarize, we observed that manipulations of stimulus frequency resulted in response bias due to the relative proportions of the stimuli (stimulus frequency, or “the signal probability effect” as in Laming, 1969), captured by changes in starting point in the model. Manipulations of decision cutoff resulted in response bias due to the subject’s perceptual adjustment to which stimuli provide strong evidence for (or against) a response, captured by changes in drift criterion in the model. Manipulations of payoff structure produced slower RTs than manipulations of stimulus frequency, in spite of the two manipulations having been set to mimic one another. In the model, this produced smaller changes in starting point due to payoff than due to stimulus frequency. Mixing manipulations of decision cutoff and of stimulus frequency resulted in a combination of decision-criterion adjustment and response bias, both identified by the model via changes in both starting point and in drift criterion.

References

- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structure. *Psychological Bulletin*, *88*, 588–606.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*, 700–765.
- Diederich, A. (2008). A further test of sequential-sampling models that account for payoff effects on response bias in perceptual decision tasks. *Perception & Psychophysics*, *70*, 229–256.
- Diederich, A., & Busemeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate change, or two-state-processing hypothesis. *Perception & Psychophysics*, *68*, 194–207.
- Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, *2*, 312–329.
- Falmagne, J.-C. (1965). Stochastic models for choice reaction time with applications to experimental results. *Journal of Mathematical Psychology*, *2*, 77–124.
- Fitts, P. M. (1966). Cognitive aspects of information process: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, *71*, 849–857.
- Geddes, J., Ratcliff, R., Allerhand, M., Childers, R., Frier, B. M., & Deary, I. J. (2010). Modeling the effects of hypoglycemia on a two-choice task. *Neuropsychology*, *24*, 652–660, doi: 10.1037/a0020074.
- Glimcher, P. W. (2005). Indeterminacy in brain and behavior. *Annual Review of Psychology*, *56*, 25–56.
- Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*, 390–394.
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574, doi: 10.1146/annurev.neuro.29.051605.113038.
- Green, D. M., & Swets, J. A. (1966). Signal detection and psychophysics. New York: Wiley.
- Halberda, J., Mazocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–669, doi: 10.1038/nature07246.
- Jarvik, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *Journal of Experimental Psychology*, *41*, 291–297.
- Kirby, N. H. (1976). Sequential effects in two-choice reaction time: Automatic facilitation or subject expectancy? *Journal of Experimental Psychology: Human perception and performance*, *2*, 567–577.
- Krajibich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*, 1292–1298, doi: 10.1038/nn.2635.
- Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*, 299–312, doi: 10.1177/1745691611406925.
- Laming, D. R. J. (1968). Information theory of choice-reaction times. New York: Academic Press.
- Laming, D. R. J. (1969). Subjective probability in choice-reaction experiments. *Journal of Mathematical Psychology*, *6*, 81–120.
- Luce, R. D. (1986). Response times: Their role in inferring elementary mental organization. New York: Oxford University Press.
- Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The Drift Diffusion Model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgment and Decision Making*, *5*, 437–449.

- Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2011). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *Manuscript submitted for publication*.
- Nelder, J. A., & Mead, R. (1965). A SIMPLEX method for function minimization. *Computer Journal*, *7*, 308–313.
- Rangel, A., Camerer, C., & Montague, R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*, 545–556, doi: 10.1038/nrn2357.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
- Ratcliff, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychological Review*, *92*, 215–225.
- Ratcliff, R. (2002). A diffusion model account of response time and accuracy in brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review*, *9*, 278–291.
- Ratcliff, R. (2008). Modeling aging effects on two-choice tasks: response signal and response time data. *Psychology and Aging*, *23*, 900–916, doi: 10.1037/a0013930.
- Ratcliff, R., Gómez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*, 159–182, doi: 10.1037/0033-295X.111.1.159.
- Ratcliff, R., Love, J., Thompson, C. A., & Opfer, J. (in press). Children are not like older adults: A diffusion model analysis of developmental changes in speeded responses. *Child Development*.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.
- Ratcliff, R., & Rouder, J. F. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human perception and performance*, *26*, 127–140.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367, doi: 10.1037/0033-295X.111.2.333.
- Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, *16*, 323–341.
- Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*, 127–157.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*, 438–481.
- Ratcliff, R., & Van Dongen, H. P. A. (2009). Sleep deprivation affects multiple distinct cognitive processes. *Psychonomic Bulletin & Review*, *16*, 742–751, doi: 10.3758/PBR.16.4.742.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *106*, 261–300.
- Remington, R. J. (1969). Analysis of sequential effects in choice reaction times. *Journal of Experimental Psychology*, *82*, 250–257.
- Rorie, A. E., Gao, J., McClelland, J. L., Newsome, W. T. (2010). Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS One*, *5*, e9308, doi:10.1371/journal.pone.0009308.
- Sanfey, A. G., Loewenstein, G., McClure, S. M., & Cohen J. D. (2006). Neuroeconomics: Cross-currents in research on decision-making. *Trends in Cognitive Sciences*, *10*, 108–116, doi: 10.1016/j.tics.2006.01.009.
- Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging*, *25*, 377–390.
- Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2005). Choosing the greater of two goods: Neural currencies for valuation and decision making. *Nature Reviews Neuroscience*, *6*, 363–375, doi: 10.1038/nrn1666.
- Swets, J. A., Tanner Jr., W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*, 301–340.
- Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*, 550–592.
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science*, *6*, 291–298, doi: 10.1177/1745691611406923.

Appendix

In Tables A1 through A4, we provide the parameter estimates for individual subjects for each condition in Experiment 1. In Tables A5 through A10, we provide the parameter estimates for each subject in Experiment 2 for each of the six models we tested. In Figures A1 through A4, we illustrate the model fits to Experiment 2 by plotting mean quantile-RT data and accuracy (across subjects) against Model II predictions obtained by using the mean parameter estimates presented in Table A6.

Table A1: Experiment 1, stimulus frequency (Condition 1).

Full model (18 free parameters)										
	a	T_{er}	s_t	z_l	z_n	z_h	s_z	d_c^l	d_c^h	
S1	0.113	0.325	0.119	0.075	0.055	0.042	0.070	-0.004	-0.017	
S2	0.147	0.377	0.150	0.087	0.072	0.056	0.057	-0.029	0.049	
S3	0.078	0.281	0.077	0.056	0.039	0.032	0.015	0.048	-0.016	
S4	0.131	0.332	0.131	0.088	0.065	0.046	0.079	-0.001	0.014	
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S1	0.148	0.516	0.412	0.252	0.100	-0.117	-0.295	-0.383	-0.492	386
S2	0.171	0.416	0.327	0.213	0.049	-0.136	-0.291	-0.428	-0.487	356
S3	0.112	0.585	0.468	0.328	0.097	-0.171	-0.353	-0.539	-0.665	546
S4	0.136	0.547	0.426	0.288	0.093	-0.125	-0.326	-0.442	-0.529	291
No- d_c model (16 free parameters)										
	a	T_{er}	s_t	z_l	z_n	z_h	s_z			
S1	0.113	0.325	0.119	0.075	0.055	0.041	0.068			
S2	0.142	0.370	0.151	0.081	0.070	0.059	0.042			
S3	0.077	0.281	0.077	0.057	0.038	0.031	0.004			
S4	0.137	0.319	0.089	0.090	0.067	0.050	0.086			
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S1	0.142	0.503	0.400	0.243	0.093	-0.122	-0.296	-0.382	-0.491	388
S2	0.144	0.372	0.298	0.186	0.036	-0.100	-0.256	-0.386	-0.437	397*
S3	0.090	0.575	0.460	0.324	0.103	-0.159	-0.327	-0.509	-0.635	555
S4	0.141	0.534	0.429	0.291	0.102	-0.105	-0.310	-0.423	-0.507	375*
Same- z model (16 free parameters)										
	a	T_{er}	s_t	z			s_z	d_c^l	d_c^h	
S1	0.124	0.327	0.147	0.064			0.098	0.136	-0.137	
S2	0.146	0.364	0.168	0.072			0.048	0.026	0.003	
S3	0.083	0.278	0.085	0.046			0.027	0.282	-0.104	
S4	0.134	0.325	0.186	0.069			0.095	0.112	-0.094	
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S1	0.249	0.658	0.508	0.314	0.110	-0.164	-0.418	-0.490	-0.667	810*
S2	0.162	0.385	0.308	0.193	0.034	-0.133	-0.287	-0.423	-0.472	537*
S3	0.211	0.633	0.514	0.370	0.093	-0.230	-0.454	-0.668	-0.797	823*
S4	0.186	0.587	0.471	0.314	0.089	-0.147	-0.370	-0.460	-0.603	771*

Note. S# = Subject number. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z_x = starting point at trials with x bias (viz., toward none (n), low (l), or high (h) responses); s_z = range of variability in starting point; d_c^x = drift criterion at the x bias condition; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup. * indicates strong support (i.e., $p > .99$) for the full model over that entry.

Table A2: Experiment 1, payoff structure (Condition 2)

Full model (18 free parameters)										
	a	T_{er}	s_t	z_l	z_n	z_h	s_z	d_c^l	d_c^h	
S5	0.175	0.404	0.180	0.089	0.084	0.085	0.004	0.006	-0.016	
S6	0.140	0.350	0.085	0.077	0.071	0.069	0.091	-0.020	-0.088	
S7	0.147	0.334	0.077	0.074	0.075	0.065	0.095	0.119	-0.022	
S8	0.084	0.411	0.130	0.038	0.034	0.033	0.004	0.000	-0.045	
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S5	0.086	0.361	0.284	0.194	0.061	-0.105	-0.219	-0.333	-0.403	541
S6	0.215	0.796	0.637	0.363	0.124	-0.117	-0.374	-0.552	-0.733	309
S7	0.165	0.672	0.508	0.316	0.084	-0.160	-0.425	-0.662	-0.776	314
S8	0.155	0.560	0.412	0.257	0.102	-0.052	-0.199	-0.342	-0.421	335
No- d_c model (16 free parameters)										
	a	T_{er}	s_t	z_l	z_n	z_h	s_z			
S5	0.162	0.402	0.153	0.083	0.078	0.075	0.023			
S6	0.138	0.348	0.088	0.077	0.072	0.064	0.084			
S7	0.145	0.330	0.082	0.079	0.071	0.060	0.081			
S8	0.084	0.412	0.131	0.039	0.035	0.031	0.015			
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S5	0.016	0.320	0.252	0.157	0.044	-0.084	-0.181	-0.288	-0.348	595*
S6	0.208	0.737	0.578	0.314	0.080	-0.148	-0.395	-0.563	-0.746	341*
S7	0.171	0.676	0.528	0.337	0.109	-0.120	-0.356	-0.585	-0.690	426*
S8	0.168	0.561	0.408	0.247	0.089	-0.069	-0.223	-0.371	-0.454	344
Same- z model (16 free parameters)										
	a	T_{er}	s_t	z			s_z	d_c^l	d_c^h	
S5	0.175	0.406	0.184	0.086			0.005	0.013	-0.022	
S6	0.135	0.353	0.097	0.068			0.088	0.002	-0.065	
S7	0.144	0.334	0.082	0.069			0.095	0.142	0.014	
S8	0.085	0.412	0.130	0.036			0.019	0.037	-0.053	
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S5	0.089	0.363	0.284	0.195	0.062	-0.106	-0.223	-0.337	-0.408	547
S6	0.204	0.784	0.590	0.365	0.085	-0.133	-0.390	-0.527	-0.703	371*
S7	0.161	0.651	0.482	0.290	0.070	-0.178	-0.430	-0.657	-0.774	416*
S8	0.170	0.570	0.418	0.255	0.095	-0.071	-0.227	-0.369	-0.452	342

Note. S# = Subject number. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z_x = starting point at trials with x bias (viz., toward none (n), low (l), or high (h) responses); s_z = range of variability in starting point; d_c^x = drift criterion at the x bias condition; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup. * indicates strong support (i.e., $p > .99$) for the full model over that entry.

Table A3: Experiment 1, Decision Cutoff (Condition 3).

Full model (18 free parameters)										
	a	T_{er}	s_t	z_{40}	z_{50}	z_{60}	s_z	d_c^{40}	d_c^{60}	
S9	0.143	0.357	0.142	0.064	0.065	0.072	0.102	-0.259	0.255	
S10	0.083	0.291	0.080	0.046	0.035	0.048	0.056	0.140	0.509	
S11	0.160	0.372	0.108	0.077	0.074	0.072	0.048	-0.130	0.085	
S12	0.129	0.356	0.152	0.060	0.059	0.062	0.091	-0.121	0.189	
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S9	0.269	0.763	0.581	0.373	0.167	-0.081	-0.353	-0.565	-0.747	428
S10	0.229	0.616	0.397	0.174	-0.035	-0.342	-0.610	-0.872	-1.061	775
S11	0.219	0.566	0.437	0.290	0.114	-0.068	-0.243	-0.345	-0.482	384
S12	0.207	0.549	0.399	0.257	0.058	-0.117	-0.307	-0.427	-0.528	444
No- d_c model (16 free parameters)										
	a	T_{er}	s_t	z_{40}	z_{50}	z_{60}	s_z			
S9	0.127	0.332	0.141	0.044	0.059	0.082	0.007			
S10	0.079	0.275	0.096	0.039	0.028	0.055	0.016			
S11	0.160	0.366	0.108	0.066	0.076	0.082	0.023			
S12	0.122	0.339	0.156	0.046	0.053	0.069	0.041			
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S9	0.198	0.471	0.338	0.204	0.058	-0.048	-0.209	-0.337	-0.485	1037*
S10	0.151	0.579	0.378	0.237	0.088	-0.068	-0.228	-0.394	-0.571	1201*
S11	0.226	0.501	0.374	0.255	0.096	-0.070	-0.237	-0.330	-0.456	588*
S12	0.182	0.436	0.309	0.211	0.059	-0.063	-0.195	-0.272	-0.390	802*
Same- z model (16 free parameters)										
	a	T_{er}	s_t	z			s_z	d_c^{40}	d_c^{60}	
S9	0.140	0.357	0.146	0.066			0.099	-0.259	0.284	
S10	0.075	0.282	0.103	0.038			0.037	0.240	0.521	
S11	0.160	0.372	0.106	0.075			0.050	-0.119	0.077	
S12	0.129	0.356	0.153	0.060			0.089	-0.112	0.202	
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S9	0.251	0.731	0.554	0.355	0.152	-0.093	-0.360	-0.566	-0.739	444*
S10	0.041	0.425	0.231	0.064	-0.123	-0.327	-0.542	-0.750	-0.898	810*
S11	0.222	0.568	0.437	0.290	0.113	-0.070	-0.243	-0.346	-0.485	389
S12	0.205	0.537	0.389	0.248	0.050	-0.122	-0.314	-0.433	-0.532	446

Note. S# = Subject number. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z_{crit} = starting point at the *crit* decision cutoff (viz., 40, 50, or 60); s_z = range of variability in starting point; d_c^{crit} = drift criterion at the *crit* decision cutoff; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup. * indicates strong support (i.e., $p > .99$) for the full model over that entry.

Table A4: Experiment 1, Decision Cutoff and Stimulus Frequency (Condition 4).

Full model (18 free parameters)										
	a	T_{er}	s_t	z_{40}	z_{50}	z_{60}	s_z	d_c^{40}	d_c^{60}	
S9	0.122	0.325	0.148	0.044	0.059	0.081	0.067	-0.183	0.208	
S10	0.080	0.272	0.103	0.026	0.042	0.054	0.025	-0.204	0.294	
S11	0.162	0.373	0.126	0.067	0.071	0.088	0.064	-0.101	0.066	
S12	0.114	0.305	0.119	0.042	0.053	0.063	0.039	-0.072	0.150	
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S9	0.107	0.490	0.362	0.250	0.068	-0.098	-0.234	-0.378	-0.496	625
S10	0.075	0.592	0.466	0.270	0.070	-0.113	-0.303	-0.488	-0.653	1218
S11	0.196	0.466	0.378	0.225	0.106	-0.094	-0.233	-0.332	-0.476	284
S12	0.039	0.246	0.214	0.120	0.017	-0.090	-0.177	-0.253	-0.323	550
No- d_c model (16 free parameters)										
	a	T_{er}	s_t	z_{40}	z_{50}	z_{60}	s_z			
S9	0.117	0.317	0.148	0.029	0.054	0.089	0.007			
S10	0.083	0.273	0.105	0.019	0.041	0.067	0.006			
S11	0.161	0.369	0.137	0.057	0.071	0.097	0.038			
S12	0.116	0.305	0.138	0.033	0.051	0.076	0.009			
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S9	0.108	0.388	0.254	0.157	0.041	-0.052	-0.133	-0.230	-0.345	1323*
S10	0.135	0.561	0.395	0.212	0.064	-0.080	-0.176	-0.359	-0.543	1936*
S11	0.199	0.438	0.341	0.208	0.084	-0.080	-0.230	-0.317	-0.454	468*
S12	0.081	0.270	0.217	0.128	0.029	-0.043	-0.121	-0.186	-0.257	1072*
Same- z model (16 free parameters)										
	a	T_{er}	s_t	z			s_z	d_c^{40}	d_c^{60}	
S9	0.120	0.319	0.175	0.060			0.075	-0.252	0.321	
S10	0.077	0.267	0.105	0.038			0.029	-0.337	0.378	
S11	0.157	0.364	0.141	0.073			0.052	-0.117	0.117	
S12	0.114	0.295	0.135	0.052			0.031	-0.113	0.197	
	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	χ^2
S9	0.097	0.495	0.369	0.255	0.060	-0.113	-0.272	-0.415	-0.534	1014*
S10	0.033	0.577	0.490	0.299	0.094	-0.095	-0.303	-0.471	-0.613	1628*
S11	0.176	0.418	0.348	0.210	0.087	-0.080	-0.230	-0.331	-0.437	365*
S12	0.048	0.243	0.218	0.124	0.015	-0.097	-0.183	-0.263	-0.329	701*

Note. S# = Subject number. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z_{crit} = starting point at the *crit* decision cutoff (viz., 40, 50, or 60); s_z = range of variability in starting point; d_c^{crit} = drift criterion at the *crit* decision cutoff; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup. * indicates strong support (i.e., $p > .99$) for the full model over that entry.

Table A5: Parameter estimates: Model I (Experiment 2).

Subj	a	T_{er}	s_t	z_l^{freq}	z_h^{freq}	z_l^{pay}	z_h^{pay}	z_{55}^{crit}	z_{45}^{crit}	
A	0.109	0.343	0.150	0.031	0.083	z_l^{freq}	z_h^{freq}	0.053	z_{55}^{crit}	
B	0.106	0.320	0.139	0.038	0.070	z_l^{freq}	z_h^{freq}	0.051	z_{55}^{crit}	
C	0.116	0.363	0.243	0.052	0.073	z_l^{freq}	z_h^{freq}	0.053	z_{55}^{crit}	
D	0.112	0.371	0.266	0.046	0.054	z_l^{freq}	z_h^{freq}	0.051	z_{55}^{crit}	
E	0.105	0.405	0.161	0.047	0.059	z_l^{freq}	z_h^{freq}	0.055	z_{55}^{crit}	
F	0.097	0.329	0.198	0.042	0.055	z_l^{freq}	z_h^{freq}	0.047	z_{55}^{crit}	
Subj	z_{55}^{cfr}	z_{45}^{cfr}	s_z^{freq}	s_z^{pay}	s_z^{crit}	s_z^{cfr}	$d_c^{b;freq}$	$d_c^{b;pay}$	$d_c^{j55;crit}$	$d_c^{j45;crit}$
A	0.040	0.078	0.023	s_z^{freq}	0.016	0.039	0	0	-0.077	0.071
B	0.038	0.070	0.068	s_z^{freq}	0.040	0.002	0	0	-0.064	0.068
C	0.052	0.073	0.079	s_z^{freq}	0.076	0.032	0	0	-0.017	0.105
D	0.047	0.055	0.089	s_z^{freq}	0.015	0.015	0	0	-0.022	0.034
E	0.047	0.055	0.024	s_z^{freq}	0.001	0.018	0	0	-0.040	0.104
F	0.045	0.053	0.066	s_z^{freq}	0.008	0.000	0	0	-0.040	0.069
Subj	$d_c^{j55;cfr}$	$d_c^{j45;cfr}$	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	χ^2
A	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.123	-0.338	-0.209	-0.068	0.082	0.228	0.328	1034.8
B	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.146	-0.399	-0.274	-0.100	0.093	0.278	0.409	1955.3
C	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.164	-0.317	-0.227	-0.098	0.049	0.183	0.290	2025.9
D	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.141	-0.330	-0.214	-0.073	0.065	0.191	0.294	1559.9
E	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.202	-0.346	-0.222	-0.096	0.067	0.205	0.324	934.8
F	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.161	-0.377	-0.288	-0.147	0.032	0.195	0.309	1404.8

Note. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z_x^{task} = starting point at trials within x condition (viz., bias toward low (l) or high (h) responses or decision cutoff set at 45 or 55) for the corresponding task manipulation (viz., stimulus frequency ($freq$), payoffs (pay), decision cutoff ($crit$), or decision cutoff and stimulus frequency (cfr); s_z^{task} = range of variability in starting point for the corresponding task; d_c^x = drift criterion at the x condition; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup.

Table A6: Parameter estimates: Model II (Experiment 2).

Subj	a	T_{er}	s_t	z_l^{freq}	z_h^{freq}	z_l^{pay}	z_h^{pay}	z_{55}^{crit}	z_{45}^{crit}	
A	0.111	0.348	0.144	0.027	0.084	0.038	0.085	0.054	z_{55}^{crit}	
B	0.107	0.326	0.131	0.022	0.085	0.055	0.056	0.052	z_{55}^{crit}	
C	0.114	0.371	0.251	0.043	0.074	0.060	0.062	0.053	z_{55}^{crit}	
D	0.113	0.375	0.269	0.040	0.052	0.054	0.058	0.052	z_{55}^{crit}	
E	0.105	0.406	0.160	0.044	0.062	0.050	0.056	0.055	z_{55}^{crit}	
F	0.099	0.328	0.192	0.032	0.061	0.049	0.054	0.048	z_{55}^{crit}	
Subj	z_{55}^{cfr}	z_{45}^{cfr}	s_z^{freq}	s_z^{pay}	s_z^{crit}	s_z^{cfr}	$d_c^{b;freq}$	$d_c^{b;pay}$	$d_c^{55;crit}$	$d_c^{45;crit}$
A	0.040	0.080	0.045	0.002	0.037	0.050	0	0	-0.079	0.082
B	0.039	0.064	0.047	0.027	0.049	0.028	0	0	-0.066	0.068
C	0.054	0.062	0.091	0.009	0.079	0.043	0	0	-0.029	0.100
D	0.047	0.057	0.089	0.090	0.011	0.016	0	0	-0.017	0.029
E	0.047	0.055	0.002	0.038	0.000	0.027	0	0	-0.041	0.101
F	0.045	0.054	0.025	0.080	0.013	0.022	0	0	-0.049	0.052
Subj	$d_c^{55;cfr}$	$d_c^{45;cfr}$	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	χ^2
A	$d_c^{55;crit}$	$d_c^{45;crit}$	0.140	-0.364	-0.227	-0.077	0.084	0.237	0.343	892.6
B	$d_c^{55;crit}$	$d_c^{45;crit}$	0.157	-0.415	-0.282	-0.099	0.092	0.279	0.418	1055.5
C	$d_c^{55;crit}$	$d_c^{45;crit}$	0.169	-0.322	-0.226	-0.092	0.058	0.198	0.307	1602.5
D	$d_c^{55;crit}$	$d_c^{45;crit}$	0.161	-0.350	-0.226	-0.079	0.073	0.200	0.301	1474.0
E	$d_c^{55;crit}$	$d_c^{45;crit}$	0.195	-0.341	-0.217	-0.092	0.064	0.200	0.318	906.1
F	$d_c^{55;crit}$	$d_c^{45;crit}$	0.165	-0.363	-0.286	-0.129	0.034	0.211	0.320	1210.5

Note. a = boundary; T_{er} = nonddecision time (in s); s_t = range of variability in T_{er} ; z_x^{task} = starting point at trials within x condition (viz., bias toward low (l) or high (h) responses or decision cutoff set at 45 or 55) for the corresponding task manipulation (viz., stimulus frequency ($freq$), payoffs (pay), decision cutoff ($crit$), or decision cutoff and stimulus frequency (cfr); s_z^{task} = range of variability in starting point for the corresponding task; d_c^x = drift criterion at the x condition; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup.

Table A7: Parameter Estimates: Model III (Experiment 2).

Subj	a	T_{er}	s_t	z_l^{freq}	z_h^{freq}	z_l^{pay}	z_h^{pay}	z_{55}^{crit}	z_{45}^{crit}	
A	0.111	0.348	0.143	0.027	0.085	0.038	0.085	0.044	0.066	
B	0.105	0.322	0.130	0.022	0.084	0.054	0.055	0.045	0.056	
C	0.115	0.371	0.251	0.043	0.074	0.059	0.062	0.054	0.055	
D	0.114	0.374	0.267	0.042	0.053	0.056	0.058	0.052	0.055	
E	0.105	0.407	0.160	0.043	0.062	0.050	0.056	0.051	0.064	
F	0.099	0.327	0.192	0.032	0.060	0.049	0.053	0.044	0.052	
Subj	z_{55}^{cfr}	z_{45}^{cfr}	s_z^{freq}	s_z^{pay}	s_z^{crit}	s_z^{cfr}	$d_c^{b;freq}$	$d_c^{b;pay}$	$d_c^{55;crit}$	$d_c^{45;crit}$
A	0.036	0.081	0.044	0.000	0.016	0.043	0	0	0	0
B	0.037	0.065	0.039	0.005	0.037	0.004	0	0	0	0
C	0.056	0.059	0.092	0.006	0.081	0.038	0	0	0	0
D	0.047	0.054	0.091	0.091	0.010	0.000	0	0	0	0
E	0.044	0.055	0.013	0.042	0.000	0.029	0	0	0	0
F	0.044	0.055	0.021	0.079	0.003	0.021	0	0	0	0
Subj	$d_c^{55;cfr}$	$d_c^{45;cfr}$	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	χ^2
A	-0.025	0.071	0.140	-0.358	-0.219	-0.076	0.076	0.225	0.330	900.1
B	-0.051	0.033	0.140	-0.378	-0.250	-0.087	0.089	0.259	0.390	1095.8
C	-0.060	0.108	0.181	-0.303	-0.203	-0.083	0.073	0.208	0.320	1696.4
D	-0.007	0.058	0.168	-0.361	-0.235	-0.087	0.071	0.193	0.300	1481.8
E	-0.006	0.102	0.200	-0.337	-0.207	-0.087	0.067	0.201	0.323	968.8
F	-0.035	0.045	0.161	-0.354	-0.278	-0.134	0.033	0.195	0.310	1260.2

Note. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z_x^{task} = starting point at trials within x condition (viz., bias toward low (l) or high (h) responses or decision cutoff set at 45 or 55) for the corresponding task manipulation (viz., stimulus frequency ($freq$), payoffs (pay), decision cutoff ($crit$), or decision cutoff and stimulus frequency (cfr); s_z^{task} = range of variability in starting point for the corresponding task; d_c^x = drift criterion at the x condition; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup.

Table A8: Parameter estimates: Model IV (Experiment 2).

Subj	a	T_{er}	s_t	z_l^{freq}	z_h^{freq}	z_l^{pay}	z_h^{pay}	z_{55}^{crit}	z_{45}^{crit}	
A	0.110	0.348	0.144	0.027	0.084	0.037	0.084	0.055	z_{55}^{crit}	
B	0.107	0.327	0.130	0.023	0.085	0.055	0.056	0.051	z_{55}^{crit}	
C	0.115	0.371	0.253	0.043	0.074	0.059	0.061	0.053	z_{55}^{crit}	
D	0.115	0.371	0.261	0.043	0.053	0.056	0.059	0.054	z_{55}^{crit}	
E	0.104	0.405	0.159	0.042	0.060	0.048	0.054	0.057	z_{55}^{crit}	
F	0.101	0.327	0.190	0.033	0.061	0.050	0.055	0.050	z_{55}^{crit}	
Subj	z_{55}^{cfr}	z_{45}^{cfr}	s_z^{freq}	s_z^{pay}	s_z^{crit}	s_z^{cfr}	$d_c^{b;freq}$	$d_c^{b;pay}$	$d_c^{55;crit}$	$d_c^{45;crit}$
A	0.033	0.085	0.044	0.000	0.033	0.030	0	0	-0.105	0.065
B	0.034	0.051	0.051	0.034	0.051	0.018	0	0	-0.058	0.093
C	0.049	0.070	0.092	0.010	0.080	0.039	0	0	-0.041	0.085
D	0.046	0.062	0.092	0.092	0.007	0.020	0	0	-0.032	0.025
E	0.042	0.061	0.005	0.038	0.003	0.011	0	0	-0.087	0.066
F	0.041	0.060	0.031	0.081	0.021	0.025	0	0	-0.071	0.060
Subj	$d_c^{55;cfr}$	$d_c^{45;cfr}$	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	χ^2
A	0	0	0.131	-0.345	-0.209	-0.066	0.083	0.234	0.341	905.4
B	0	0	0.156	-0.416	-0.282	-0.101	0.090	0.274	0.414	1070.8
C	0	0	0.175	-0.309	-0.211	-0.080	0.075	0.212	0.319	1685.0
D	0	0	0.156	-0.350	-0.219	-0.076	0.066	0.196	0.295	1496.9
E	0	0	0.186	-0.312	-0.187	-0.070	0.082	0.214	0.330	946.7
F	0	0	0.172	-0.367	-0.289	-0.142	0.041	0.212	0.328	1232.0

Note. a = boundary; T_{er} = nondecision time (in s); s_t = range of variability in T_{er} ; z_x^{task} = starting point at trials within x condition (viz., bias toward low (l) or high (h) responses or decision cutoff set at 45 or 55) for the corresponding task manipulation (viz., stimulus frequency ($freq$), payoffs (pay), decision cutoff ($crit$), or decision cutoff and stimulus frequency (cfr); s_z^{task} = range of variability in starting point for the corresponding task; d_c^x = drift criterion at the x condition; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup.

Table A9: Parameter estimates: Model V (Experiment 2).

Subj	a	T_{er}	s_t	z_l^{freq}	z_h^{freq}	z_l^{pay}	z_h^{pay}	z_{55}^{crit}	z_{45}^{crit}	
A	0.112	0.347	0.143	0.028	0.085	0.038	0.086	0.055	z_{55}^{crit}	
B	0.105	0.321	0.131	0.022	0.083	0.054	0.055	0.051	z_{55}^{crit}	
C	0.116	0.367	0.246	0.045	0.076	0.061	0.064	0.053	z_{55}^{crit}	
D	0.112	0.375	0.269	0.041	0.052	0.053	0.056	0.051	z_{55}^{crit}	
E	0.105	0.405	0.162	0.044	0.062	0.050	0.056	0.055	z_{55}^{crit}	
F	0.097	0.328	0.194	0.031	0.058	0.047	0.052	0.046	z_{55}^{crit}	
Subj	z_{55}^{cfr}	z_{45}^{cfr}	s_z^{freq}	s_z^{pay}	s_z^{crit}	s_z^{cfr}	$d_c^{b;freq}$	$d_c^{b;pay}$	$d_c^{j55;crit}$	$d_c^{j45;crit}$
A	0.062	z_{55}^{cfr}	0.044	0.003	0.033	0.090	0	0	-0.131	0.123
B	0.051	z_{55}^{cfr}	0.037	0.004	0.040	0.045	0	0	-0.095	0.102
C	0.059	z_{55}^{cfr}	0.092	0.014	0.079	0.040	0	0	-0.036	0.109
D	0.051	z_{55}^{cfr}	0.090	0.090	0.010	0.009	0	0	-0.045	0.030
E	0.050	z_{55}^{cfr}	0.002	0.037	0.000	0.030	0	0	-0.051	0.113
F	0.048	z_{55}^{cfr}	0.024	0.075	0.008	0.025	0	0	-0.062	0.062
Subj	$d_c^{j55;cfr}$	$d_c^{j45;cfr}$	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	χ^2
A	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.153	-0.383	-0.239	-0.074	0.091	0.254	0.360	1305.6
B	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.139	-0.395	-0.273	-0.096	0.084	0.268	0.390	1231.6
C	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.169	-0.323	-0.229	-0.097	0.054	0.194	0.301	1610.9
D	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.162	-0.338	-0.215	-0.066	0.078	0.207	0.320	1496.4
E	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.198	-0.344	-0.220	-0.094	0.066	0.203	0.320	925.5
F	$d_c^{j55;crit}$	$d_c^{j45;crit}$	0.143	-0.339	-0.268	-0.130	0.041	0.203	0.319	1229.4

Note. a = boundary; T_{er} = nonddecision time (in s); s_t = range of variability in T_{er} ; z_x^{task} = starting point at trials within x condition (viz., bias toward low (l) or high (h) responses or decision cutoff set at 45 or 55) for the corresponding task manipulation (viz., stimulus frequency ($freq$), payoffs (pay), decision cutoff ($crit$), or decision cutoff and stimulus frequency (cfr); s_z^{task} = range of variability in starting point for the corresponding task; d_c^x = drift criterion at the x condition; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup.

Table A10: Parameter estimates: Model VI (Experiment 2).

Subj	a	T_{er}	s_t	z_l^{freq}	z_h^{freq}	z_l^{pay}	z_h^{pay}	z_{55}^{crit}	z_{45}^{crit}	
A	0.113	0.336	0.143	0.056	z_l^{freq}	0.063	z_l^{pay}	0.056	z_{55}^{crit}	
B	0.106	0.310	0.145	0.050	z_l^{freq}	0.059	z_l^{pay}	0.052	z_{55}^{crit}	
C	0.118	0.369	0.237	0.060	z_l^{freq}	0.067	z_l^{pay}	0.053	z_{55}^{crit}	
D	0.115	0.372	0.268	0.045	z_l^{freq}	0.057	z_l^{pay}	0.053	z_{55}^{crit}	
E	0.107	0.404	0.163	0.052	z_l^{freq}	0.054	z_l^{pay}	0.056	z_{55}^{crit}	
F	0.098	0.330	0.199	0.042	z_l^{freq}	0.053	z_l^{pay}	0.047	z_{55}^{crit}	
Subj	z_{55}^{cfr}	z_{45}^{cfr}	s_z^{freq}	s_z^{pay}	s_z^{crit}	s_z^{cfr}	$d_c^{b;freq}$	$d_c^{b;pay}$	$d_c^{55;crit}$	$d_c^{45;crit}$
A	0.043	0.080	0.085	0.032	0.000	0.033	0.355	0.358	0.055	0.247
B	0.041	0.063	0.085	0.000	0.017	0.001	0.221	0.037	-0.015	0.122
C	0.056	0.062	0.094	0.006	0.085	0.045	0.271	0.038	0.026	0.169
D	0.050	0.058	0.085	0.091	0.019	0.022	0.065	0.022	-0.025	0.043
E	0.048	0.056	0.020	0.038	0.000	0.013	0.133	0.093	0.004	0.153
F	0.044	0.053	0.042	0.076	0.012	0.012	0.150	0.012	-0.001	0.109
Subj	$d_c^{55;cfr}$	$d_c^{45;cfr}$	η	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	χ^2
A	$d_c^{55;crit}$	$d_c^{45;crit}$	0.176	-0.552	-0.405	-0.241	-0.068	0.117	0.220	1461.4
B	$d_c^{55;crit}$	$d_c^{45;crit}$	0.139	-0.448	-0.323	-0.150	0.029	0.213	0.328	2109.6
C	$d_c^{55;crit}$	$d_c^{45;crit}$	0.198	-0.398	-0.296	-0.153	-0.005	0.153	0.281	1721.6
D	$d_c^{55;crit}$	$d_c^{45;crit}$	0.177	-0.374	-0.247	-0.091	0.063	0.196	0.307	1570.4
E	$d_c^{55;crit}$	$d_c^{45;crit}$	0.215	-0.403	-0.275	-0.144	0.018	0.165	0.288	905.4
F	$d_c^{55;crit}$	$d_c^{45;crit}$	0.175	-0.423	-0.340	-0.188	-0.011	0.163	0.280	1359.8

Note. a = boundary; T_{er} = nonddecision time (in s); s_t = range of variability in T_{er} ; z_x^{task} = starting point at trials within x condition (viz., bias toward low (l) or high (h) responses or decision cutoff set at 45 or 55) for the corresponding task manipulation (viz., stimulus frequency ($freq$), payoffs (pay), decision cutoff ($crit$), or decision cutoff and stimulus frequency (cfr)); s_z^{task} = range of variability in starting point for the corresponding task; d_c^x = drift criterion at the x condition; η = variability in drift rate; ν_n = drift rate for the n stimulus subgroup.

Figure A1: Quantile-Probability plots for data and Model II predictions from Condition 1. Quantile-RT data points, averaged across subjects, are plotted in ascending order, from .1 to .9 in each column of circles. In each panel, reading it from left to right, there are three such columns across error responses followed by three columns across correct responses, making up the six difficulty conditions (for asterisk counts of 36–40, 41–45, 46–50, 51–55, 56–60, and 61–65). The horizontal position at which each quantile-RT column is plotted is the response proportion corresponding to that difficulty condition. Model predictions are plotted as crosses, connected at the same quantile levels across difficulty conditions. Note that error quantiles in conditions with fewer than five observations (for each subject) could not be computed. For error columns with eleven or fewer data points, only the median RT was plotted (excluding subjects with no error responses) to indicate the level of accuracy in those conditions (as a diamond).

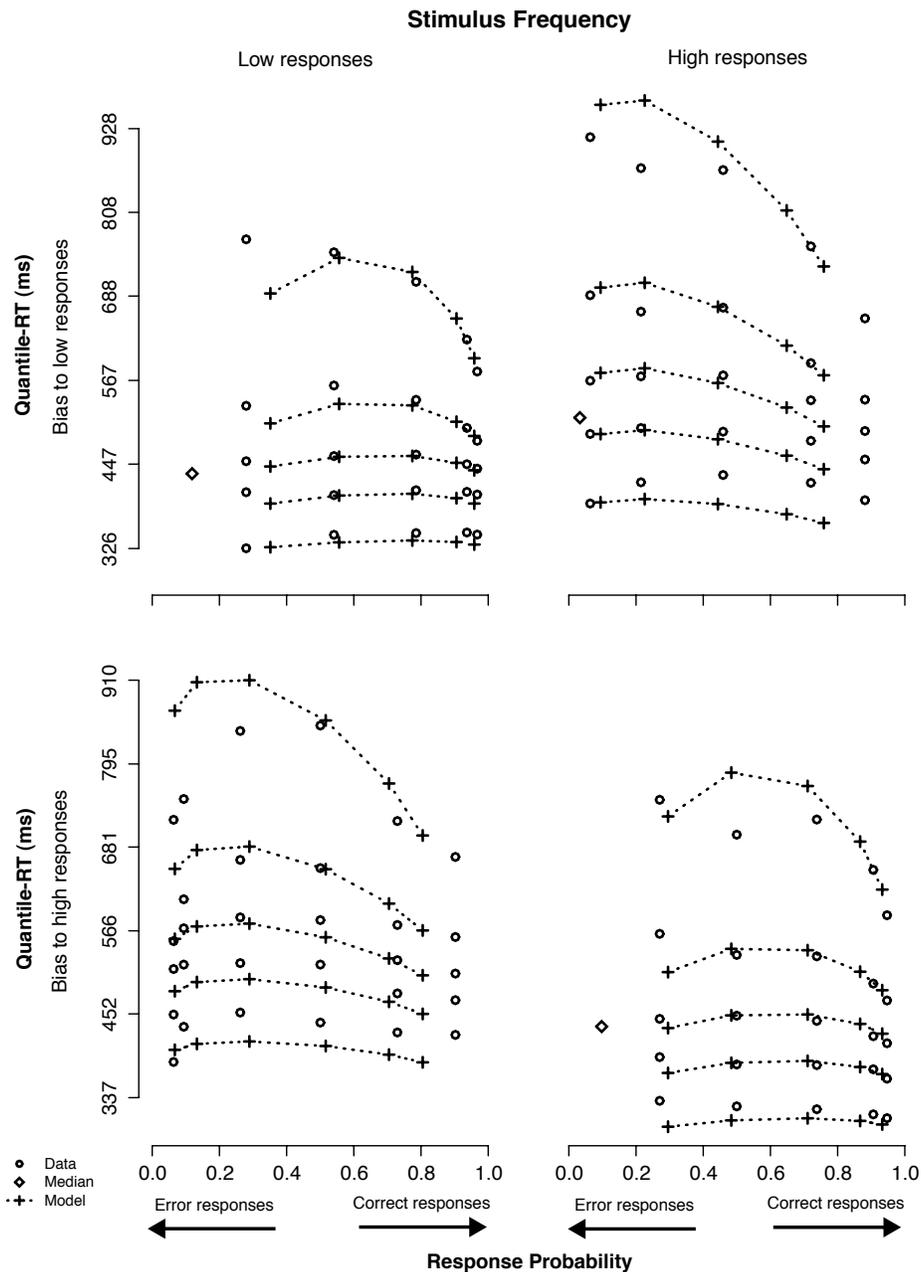


Figure A3: Quantile-Probability plots for data and Model II predictions from Condition 3. Quantile-RT data points, averaged across subjects, are plotted in ascending order, from .1 to .9 in each column of circles. In each panel, reading it from left to right, there are three such columns across error responses followed by three columns across correct responses, making up the six difficulty conditions (for asterisk counts of 36–40, 41–45, 46–50, 51–55, 56–60, and 61–65). The horizontal position at which each quantile-RT column is plotted is the response proportion corresponding to that difficulty condition. Model predictions are plotted as crosses, connected at the same quantile levels across difficulty conditions. Note that error quantiles in conditions with fewer than five observations (for each subject) could not be computed. For error columns with eleven or fewer data points, only the median RT was plotted (excluding subjects with no error responses) to indicate the level of accuracy in those conditions (as a diamond). Discontinuation of dotted lines emphasize the separation between correct and error responses.

