RICE UNIVERSITY

Factors Influencing Speed-Accuracy Tradeoffs in Decision Making

by

Jeffrey C. Zemla

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Michael D. Byrne, Ph. D. Professor,
Psychology and Computer Science

Philip T. Kortum, Ph. D. Assistant
Professor, Psychology

Tatiana Schnur, Ph. D. Assistant
Professor, Psychology

Richard R. Batsell, Ph. D. Associate
Professor, Marketing

HOUSTON, TEXAS
September 2014

Abstract

Factors Influencing Speed-Accuracy Tradeoffs in Decision Making

by

Jeffrey C. Zemla

Many simple decisions allow us to trade off between speed and accuracy. When time is critical, decisions can be made quickly but accuracy suffers. Conversely, one may spend more time making a decision which often results in more accurate decisions. Speed-accuracy tradeoffs have been studied in a number of domains including motor control (Fitts, 1954), perception (Usher & McClelland, 2001), and higher order reasoning (Kahneman & Frederick, 2002).

Recent research has examined a set of normative models for how one should trade off speed and accuracy (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006); that is, how long someone should spend deliberating prior to action in order to maximize some reward. However, empirical work has shown haphazard adherence to these normative models (e.g., Zacksenhouse, Bogacz, & Holmes, 2010). While some subjects behave optimally, many do not.

In two experiments, several factors that affect speed-accuracy tradeoffs in a perceptual decision-making task are investigated. In one experiment, it was found that feedback and shorter blocks not only improved participants' task ability, but also resulted in more optimal speed-accuracy tradeoffs. In a second experiment, manipulating trial difficulty and subjects' awareness of difficulty level affected task performance. However, despite predictions from a normative theory, participants did not engage in an optimal speed-accuracy tradeoff policy.

Contents

List of Tables

List of Figures

Introduction

Consider the following situation: a student is taking a multiple-choice exam. There are only two minutes remaining, but he has four unanswered questions. He spends a few seconds looking at one question, but none of the choices seem familiar, so he quickly makes a guess and moves on. The choices to the next question look very familiar—he remembers seeing the answer in his textbook, but cannot recall what it is. Still, he thinks that if he stares at the question long enough the answer will come to him. In the last few seconds, the answer suddenly pops into his head and he marks it down—but now time is up and he has left two questions blank.

The decision to spend more time on some problems than others can have important consequences, but how is this decision made? An optimal solution to this problem is complex, and requires several pieces of information. Important questions to consider include: How difficult is each problem? If more time is spent on a problem, do the chances of getting it correct increase? How will the time spent on one problem affect the amount of time left for other problems?

Over the past few decades, a psychological framework has emerged to describe how people are able to trade off speed and accuracy in decision making (Ratcliff, 1978; Ratcliff & Rouder, 1998). This framework serves to describe the process by which humans make simple decisions. It is only recently, however, that researchers have started using this framework to assess the optimality of these tradeoffs with respect to normative theories (Bogacz, Hu, Holmes, & Cohen, 2010; Bogacz et al., 2006; Simen et al., 2009; Zacksenhouse et al., 2010; Balci et al., 2011). That is, while it has long been known that humans are capable of trading off speed and accuracy in decision making, it is still unclear whether they are able to make *optimal* tradeoffs, or whether they exhibit systematic biases in their decision making.

In two experiments, I explore factors that influence a person's ability to trade off speed and accuracy in a decision making task. Do people manage their time in such a way to maximize reward? If not, what causes them to behave sub-optimally? What cues help people determine how much time they should spend on any given problem?

Several factors that may affect a person's ability to make fast and accurate decisions are examined. These factors include feedback, block length, foreknowledge of task difficulty, and variability of trial difficulty. These manipulations provide cues that may influence a person's ability to manage their time effectively. Previous research has shown that a minority (roughly 30%) of subjects are able to effectively trade off speed and accuracy in a simple perceptual decision making task (Zacksenhouse et al., 2010). It is hoped that providing additional cues may assist people in selecting a more appropriate speed-accuracy tradeoff policy.

Background

Sequential Sampling Models

Over the course of many years, a broad class of models known as *sequential sampling models* have emerged as a comprehensive solution to the question of how humans are able to make simple binary decisions (Ratcliff, 1978; Ratcliff & Rouder, 1998; Usher & McClelland, 2001; Brown & Heathcote, 2008; Busemeyer & Townsend, 1993; Wagenmakers, Van Der Maas, & Grasman, 2007). Sequential sampling models have been used extensively to model signal detection tasks: a subject is presented with an ambiguous stimulus that is drawn from one of two underlying distributions, and the subject must determine which class the stimulus is a member of. Examples of signal detection tasks include perceptual tasks such as brightness discrimination (dark or light) or length discrimination (short or long). A stimulus in these experiments offers clues as to which class it belongs, but noise in the environment and in our perceptual abilities make classification ambiguous. Sequential sampling models propose that we are able to make judgements about a stimulus by repeatedly sampling the stimulus and integrating evidence over time; that is, our brain repeatedly computes class membership of a stimulus until some level of confidence is

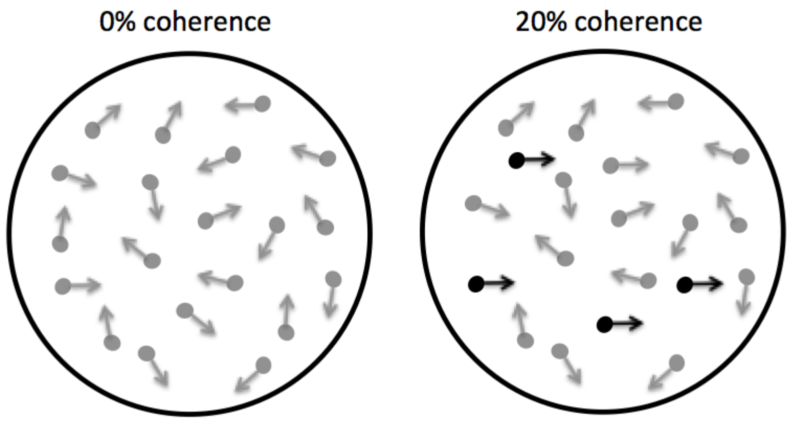

Figure 1. In the random dot motion paradigm, a cluster of dots move in random directions. The task is to decide which direction, on average, the dots are moving (left or right). The proportion of dots moving in a predetermined direction, known as the coherence level, affects the difficulty of the task. In this figure, arrows and shading are for expository purposes only.

reached. The signal-to-noise ratio of a stimulus determines how quickly and accurately it will be perceived.

To illustrate this, consider the random dot motion paradigm (Britten, Shadlen, Newsome, & Movshon, 1992) in which a subject is presented with a cluster of moving dots (see Figure 1). The majority of the dots move in a random direction, though a small proportion of them move coherently to the left or the right. The task of the subject is to decide whether the prevailing motion of the dots is to the left or the right, and respond accordingly with either a button press or a saccade towards the direction of motion. Shadlen and Newsome (1996, 2001) conducted this experiment with monkeys and found that in the middle temporal area (MT), an area of the brain associated with motion, neurons fired selectively for dots moving in a particular direction. This is consistent with previous research showing that the firing rate of neurons in MT roughly represents the instantaneous direction of motion of a stimulus (Zeki, 1974). Neurons in the lateral intraparietal area (LIP), on the other hand, were found to correlate with the integration of neuronal firing in MT, and thus reflect the average direction of motion over a short time span.

In this simplified example, the firing rate of neurons in MT represents the available evidence being sampled, whereas the firing rate of neurons in LIP represents the accumulation of evidence over time. Thus the firing rate of neurons in the LIP may represent the decision variable which is directly responsible for a subject's response. As shown in Figure 2, the firing rate of neurons in LIP is highly predictive of behavioral measures such as response time and accuracy.

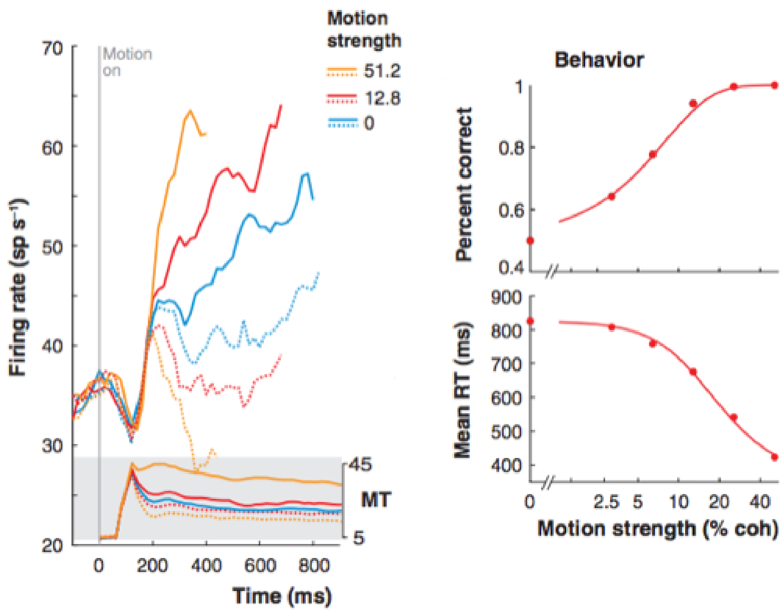

Figure 2. Neurons in MT (left, bottom) fire relative to the motion strength, i.e., the proportion of dots moving left or right. The firing rate of these neurons is constant over time as long as the motion strength remains fixed. Neurons in LIP (left, top) represent an integration of the firing rate of MT neurons over time. Thus in MT, the difference in firing rate of left-selective and right-selective neurons remains relatively fixed over time, whereas the corresponding firing rates of left- and right-selective neurons in LIP diverge over time. The pattern of firing rates in LIP is highly predictive of both behavioral response time and accuracy (right). Figure from Gold and Shadlen (2007).

The precise mathematical basis by which evidence is accumulated is subject to considerable debate. For instance, some models (Vickers, 1970; Smith & Vickers, 1988) hypothesize separate accumulators for each response option, whereas others (Ratcliff, 1978) model accumulation of evidence using a single binary decision variable. Additional debates concern whether the amount of evidence needed is fixed or relative, whether there is inhibition between response accumulators, or whether evidence decays over time (Usher & McClelland, 2001). While these differences may highlight important conceptual issues, they often have a relatively small impact on behavioral predictions.

Though sequential sampling models are often used to model signal detection tasks, they have been used to model a wide wide variety of popular psychological paradigms, including the random dot motion task (Simen et al., 2009), the lexical decision task (Ratcliff, Gomez, & McKoon, 2004), the picture word interference task (van Maanen, van Rijn, & Borst, 2009), the implicit association test (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007), and the go/no-go task (Shenoy & Yu, 2012), among others. Though these tasks are generally employed for very different purposes, sequential sampling models have proven successful in modeling the decision process underlying all of them. Thus, sequential sampling models describe a generalized process used to make simple two-alternative forced choice tasks that are widely prevalent in psychology.

Moreover, sequential sampling models have been successful in explaining decision making at multiple levels (Marr, 1982). Sequential sampling models are typically described as process models, sometimes being implemented as neural networks (Usher & McClelland, 2001) or within cognitive architectures (Lewis, Shvartsman, & Singh, 2013; van Maanen, van Rijn, & Taatgen, 2012), though there is a growing body of evidence that such a process is implemented in the brain (Gold & Shadlen, 2007; Bogacz, 2007) and that this process is rationally adapted to the environment (Bogacz et al., 2006).

Ratcliff Diffusion Model

One of the most successful and widely used sequential sampling models is the Ratcliff diffusion model (RDM; Ratcliff, 1978; Ratcliff & Rouder, 1998). The RDM posits that the decision making process can be described as a Wiener diffusion process with drift (Feller, 1968), shown in Figure 3.

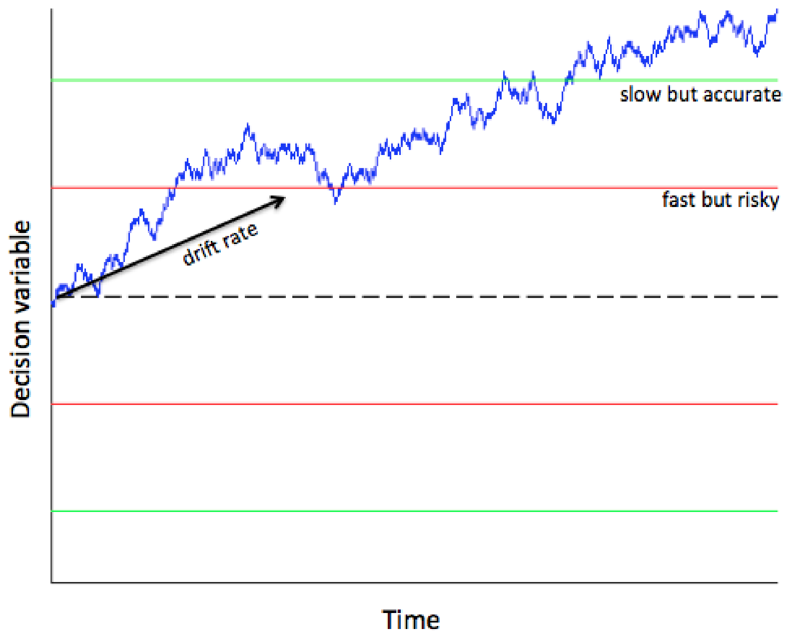

Figure 3. The Wiener diffusion process describes a stochastic process in which the value of a decision variable (blue) drifts towards an upper or lower boundary. The upper and lower boundaries (red or green) represent binary response options. Here, the green boundary represents a conservative threshold: the process will take longer to complete, but the outcome is more stable. Conversely, the red boundary will result in a shorter decision time, but the outcome is more susceptible to noise.

The simple version of the RDM (Ratcliff, 1978) contains three free parameters: a drift rate (v), a decision threshold (a), and a non-decision component (T_{er}). The drift rate reflects the average rate of evidence accumulation over time. It is analogous to sensitivity (d') in signal detection theory (Green, Swets, et al., 1966), and represents a subject's ability to discriminate between two signal classes. Thus, the drift rate is a function of both the stimulus itself and the subject's ability to perceive that stimulus. High drift rate values reflect the ability of a subject to discern a stimulus class accurately.

Table 1

Ratcliff Diffusion Model Parameters

Parameter	Description
a	Decision threshold
σ	Variability in drift rate within trial (scaling parameter)
z	Mean starting point
v	Mean drift rate
T_{er}	Mean non-decision time
s_z	Variability in starting point
s_v	Variability in drift rate across trials
s_{ter}	Variability in non-decision time

The decision threshold denotes the amount of evidence required by the subject before making a response. A high decision threshold means that the subject requires a lot of evidence (high confidence) before making a response; a low decision threshold means the subject is willing to respond with less evidence. Though choosing a high decision threshold will lead to more accurate responses, it also necessitates that those responses will be slower as more time is needed to accumulate evidence. Thus, the decision threshold directly controls the speed-accuracy tradeoff policy of a subject.

Lastly, the non-decision time accounts for any time in the process that is not strictly attributable to decision making. This may incorporate time for stimulus encoding or motor response time. Additionally, there is a single non-free parameter (σ) which reflects the amount of intrinsic noise in the process. This parameter is often referred to as a scaling parameter, as all other parameters can be described by their proportion to σ. By convention, this parameter is set to 0.1.

The RDM has been extended to include additional parameters (Ratcliff & Rouder, 1998). A biased starting point (z) can be used to model a subject's prior belief that the two outcomes are not equally likely, or an imbalance in the value of the two outcomes. This parameter is analogous to beta (β) in signal detection theory. Between-trial variability in drift rate (η), thought to reflect variability in stimulus encoding, is necessary to account for slower response times for incorrect compared to correct responses. Conversely, variability in starting point (s_z) is used to model fast errors, perhaps due to residual activation from previous trials. Lastly, variability in non-decision time (s_{ter}) is used to account for a shift in the leading edge of the response distribution (Ratcliff, Thapar, & McKoon, 2004), and reflects variability in encoding and motor response processes.

While the full parameter set (Table 1) is needed to model many nuanced behavioral phenomenon, the simple three-parameter model is frequently used for mathematical tractability and clarity. Predictions from the reduced three-parameter model correlate highly with the more flexible but higher-parameter models (Bogacz et al., 2010), and thus is useful when one is not interested in modeling these nuanced effects.

The RDM is notable in that it predicts not just error rate and average response time, but instead predicts entire distributions of response times. Additionally, it can predict both slow and fast errors, independent response time distributions for both correct and incorrect responses, and account for the interactions between speed and accuracy.

In the Ratcliff Diffusion Model, the decision threshold is under the direct volitional control of the subject. A critical question is how participants choose an appropriate decision threshold. Intuitively, a subject should set a high threshold when accuracy is important, and a low threshold when speed is important. Indeed, when subjects are alternately instructed to respond quickly and accurately, subjects are capable of complying, and responses can be modeled solely as a difference in threshold level (Ratcliff & Rouder, 1998).

However it is not immediately clear how one should choose a threshold in the absence of explicit instruction. It has been suggested that participants strive to minimize Bayes Risk, a weighted combination of accuracy and decision time (Busemeyer & Rapoport, 1988). This definition may not be particularly useful though, as it simply shifts the question to how these terms should be weighted.

If, however, the relative values of accuracy and time are made more explicit, a normative solution may be present. Consider a common scenario, in which a subject must maximize a task score in a finite amount of time. Such scenarios are common in the real world, for instance when a student seeks to maximize his score on a time test. In this case, the normative solution is to maximize the *reward rate*; that is, to maximize one's score per unit time. This idea was recently proposed by Bogacz et al. (2006), and is mathematically similar to a popular normative theory in ecology: optimal foraging theory.

Optimal Foraging Theory

Optimal foraging theory (OFT) is a prominent theory in ecology designed to test the hypothesis that animal foraging and predatory behavior is an optimal response to the organism's environment (Davies, Krebs, & West, 2012; Stephens & Krebs, 1987). A central component of OFT involves determining an animal's response to a diminishing food supply: At what point should an animal abandon a low-yield resource in favor of seeking a more plentiful resource?

Suppose an animal forages in a patch of berries. As it consumes the berries in the patch, it begins to take longer to find more berries within the patch. Thus, the rate of return (berries consumed per unit time) diminishes as the animal exploits the resource. The rate of return is described as:

$$R = \frac{G(t_W)}{t_B + t_W} \tag{1}$$

where R is the global rate of return, t_W is the time spent foraging within a patch, t_B

is the time spent between patches (i.e., searching for a new patch), and $G(t_W)$ is the gain (e.g., total number of berries) achieved from leaving a patch at time t_W. This equation describes the relation between the total gain achieved and the time needed to acquire that gain.

A foraging animal faces a choice: at any point, it can continue to search for berries within a patch, or it can leave to seek a more plentiful patch. Charnov (1976) posits that the normative response in such a situation is to exploit a patch until the rate of return from the patch drops below the average rate of return across many patches. This idea is embodied in Charnov's *marginal value theorem*:

$$\frac{d}{dt_W^*}G(t_W^*) = R(t_W^*) \tag{2}$$

or equivalently:

$$\frac{d}{dt_W^*}R(t_W^*) = 0 \tag{3}$$

If $G(t_W)$ increases asymptotically there exists a unique positive solution t_W^* to this equation, illustrated in Figure 4. $G(t)$ plots a gain curve of diminishing returns. By convention, between patch time (t_B) is plotted to the left of the y-axis, and within patch time (t_W) is plotted to the right of the y-axis. A line can be drawn from the time of departure from the previous patch to the point on $G(t)$ where an organism departs from the new patch. The slope of the line is equal to the rate of return, R. In order to maximize R, the optimal solution is to choose a t_W such that R lies tangent to $G(t)$ (satisfying Equation 2), as in R_1. If an organism departs the patch later (t_2) or earlier (t_3) than t_1^*, the overall rate of return for the organism is less than optimal. As such, the optimal time spent foraging within a patch depends not only on the shape of the gain curve, but also on the time spent between patches.

This solution describes the normative response when $G(t)$ remains the same across all patches. However the framework can be extended to accommodate multiple patch types,

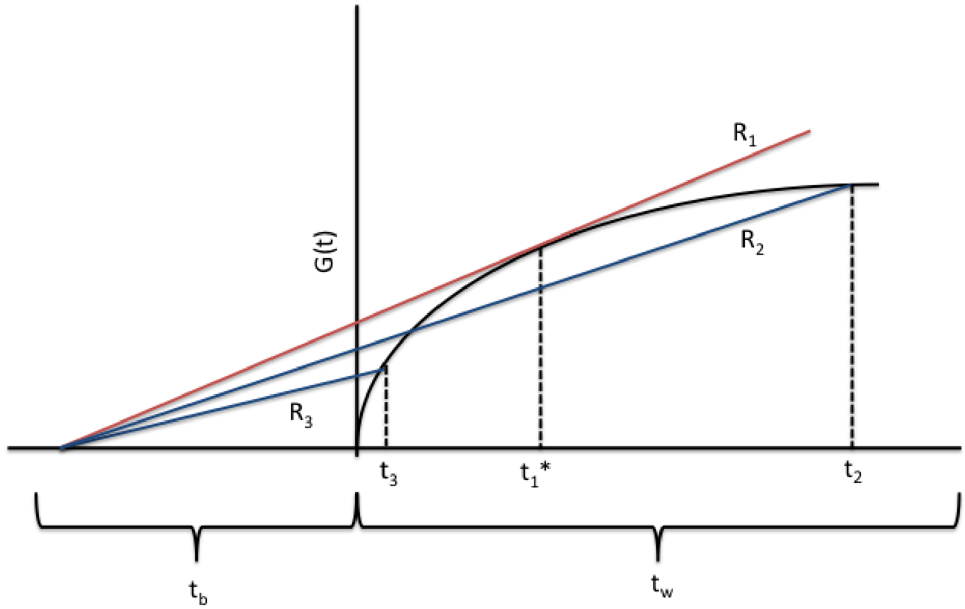

Figure 4. The solid black curve indicates a gain function, with time on the x-axis and cumulative gain on the y-axis. Rate of return is maximized (red line) when the derivative of the gain curve at t_w is equal to the rate of return.

i.e., when some patches are more profitable than others:

$$R = \frac{\sum_{i=1}^{N} \lambda_i G_i(t_{Wi})}{t_B + \sum_{i=1}^{N} \lambda_i t_{Wi}} \tag{4}$$

Here, N denotes the number of patch types, G_i denotes the gain curve of patch type i, and t_{Wi} denote the time spent within patch type i. λ_i denotes the probability of encountering patch type i, where $\sum_{i=1}^{N} \lambda_i = 1$. In this case, each patch has its own optimal within-patch time t_{Wi}. Again, a unique solution exists to this problem, and is given by:

$$
\begin{aligned}
\frac{\partial}{\partial t^*_{W1}} R(t^*_{W1}, t^*_{W2}, ..., t^*_{WN}) &= 0 \\
\frac{\partial}{\partial t^*_{W2}} R(t^*_{W1}, t^*_{W2}, ..., t^*_{WN}) &= 0 \\
&... \\
\frac{\partial}{\partial t^*_{WN}} R(t^*_{W1}, t^*_{W2}, ..., t^*_{WN}) &= 0
\end{aligned} \tag{5}
$$

Thus for a finite set of patch types, a vector of unique optimal departure times exists. The optimal values depend on the time between patches, the shape of each patch's gain curve, and the prevalence of each patch type.

Although optimal foraging theory was developed within the field of ecology, psychologists have recently began using the theory to understand cognitive processes. OFT has been used to model domains such as human information seeking on the web (Pirolli, 2007), semantic memory retrieval (Hills, Jones, & Todd, 2012), and visual search (Cain, Vul, Clark, & Mitroff, 2012).

Optimal Threshold Setting

Bogacz et al. (2006) hypothesized that decision makers should trade off speed and accuracy in a manner analogous to optimal foraging theory. Equation 1 can be expressed in terms of the Ratcliff diffusion model. In this case, $G(v, a)$ represents task accuracy as a function of drift rate and decision threshold (Ratcliff & Tuerlinckx, 2002; Bogacz et al., 2010):

$$
G(v, a) = 1 - \frac{1}{1 + e^{\frac{2va}{\sigma^2}}} \tag{6}
$$

t_W denotes the average decision time as a function of the decision threshold:

$$t_W = \frac{a}{v} tanh(\frac{va}{\sigma^2}) + T_{er} \tag{7}$$

t_B denotes the time between trials (inter-stimulus interval).

To determine the optimal thresholds, one must first fit the diffusion model to behavioral data to determine the drift rate (v) and non-decision time (T_{er}) for a given subject and task. These values may be substituted in to Equation 6 and Equation 7, leaving a single unknown variable (decision threshold, a). Note that these equations hold only in the restricted three-parameter RDM; the extended parameter set does not allow for an analytical expression of accuracy and decision time.

These predictions were tested in a series of experiments, reviewed in Holmes and Cohen (2014). Simen et al. (2009) examined several factors that are thought to influence selection of a decision threshold, including manipulation of the inter-stimulus interval (ISI or t_B). In a random dot motion experiment, they found that changes in the ISI across blocks (.5, 1, or 2 seconds) led participants to adopt different decision thresholds. Specifically, participants chose lower thresholds when the ISI was shorter, in line with predictions of reward rate maximization. Critically, manipulating the ISI should not affect one's decision threshold in several other normative theories, such as minimization of Bayes Risk or decision time, or maximization of accuracy. Additional experiments found that biasing the stimulus probability (i.e., making one direction of motion more probable than another) or manipulating the relative reward of these responses led to a shift in the RDM's starting point towards the more probable or more valuable stimulus. This finding is independent from decision threshold selection, but still supports the theory of reward rate maximization put forth by Bogacz et al. (2006).

One limitation of this finding is the authors' reliance on pooling subject data together, a technique which risks misrepresenting the true behavior of individual subjects (Estes & Maddox, 2005) in exchange for statistical power. Additionally, the authors found that participants tended to choose a threshold higher than optimal, and this deviation grew as the ISI was reduced. Thus while these findings support some of the general principles behind the reward rate maximization theory, a more direct comparison between empirical results and theoretical predictions demonstrated systematic violations of a strict adherence to the normative theory.

Bogacz et al. (2010) corroborated these findings by showing that participants modulate their decision threshold based on the ISI as well as time penalties for incorrect responses. These manipulations, which both extended the average trial time, led participants to adopt a more conservative threshold as predicted by the theory of reward rate maximization. Again, however, a more direct test of the theory comparing optimal to observed thresholds shed some doubt on the theory. Most participants chose decision thresholds that were higher than optimal, supporting an alternative theory that participants are particularly sensitive to accuracy, and maximize a weighted combination of reward rate and accuracy. The authors also tested a third hypothesis, that participants implicitly assume a weighted penalty for incorrect responses, but ultimately found that this theory did not fit the data quite as well as the weighted combination of reward rate and accuracy.

Balci et al. (2011) re-examined previous findings with an extended version of the task. They found that while initially subjects had a bias towards accuracy, as shown previously, participant thresholds approached optimality with extended practice on the task. This suggests that decision makers act conservatively as they learn a new task, but after they become more familiar with a task they are able to perform near optimally. This is perhaps the most direct evidence in support of the reward rate maximization theory.

The authors also examined how participants perform when the difficulty of the task is varied across blocks. An ideal participant should choose a unique threshold for each block, as trial difficulty affects the shape of the speed-accuracy curve and thus the optimal threshold. Instead, it was found that participants chose a single threshold that performed well across blocks; that is, they chose a threshold that worked best for the average difficulty across blocks. Note that despite the manipulation in trial difficulty, an optimal solution does not entail the multi-patch solution given in Equation 4 and Equation 5. This is because within any block, every trial is exactly the same difficulty. As such, participants cannot trade off between trial types. For instance, they cannot answer more quickly on harder trials in the hopes that the next trial will be easier.

It is unclear why participants seemed to choose a single threshold across blocks, though explanations abound. It may reflect some limit in the ability of participants to rapidly shift decision thresholds because it is computationally difficult, or perhaps switching thresholds incurs a cost of cognitive control (Holmes & Cohen, 2014). Alternatively, it may be a more pragmatic choice to choose a robust strategy that does not rely upon one's ability to discern differences in difficulty (trial type), which in itself may be a difficult task.

Zacksenhouse et al. (2010) found that roughly 30% of participants perform according to an optimal strategy. However, the remaining 70% of the data can be explained as a tradeoff between optimality and robustness. This theory assumes that there is uncertainty in the timing of the experiment; that is, participants may not be able to accurately estimate the delay between trials. When uncertainty is taken into account, participants seem to choose a robust strategy such that the chosen decision threshold will result in decent performance for a range of inter-stimulus intervals. Specifically, participants may choose a threshold that maximizes the reward for the worst possible scenario in a given range of inter-stimulus intervals.

Overall it seems that while participants are sensitive to reward rate, behavior deviates systematically from that predicted by the reward rate maximization theory. Participants almost always choose a threshold that is too high, though with extended practice on a task the threshold becomes closer to optimal. This initial tendency to set a threshold too high might be explained by a preference for accuracy (over reward) or noise in a subject's ability to estimate inter-stimulus intervals.

Current Research

Overview

Previous research has shown that under some circumstances, in the absence of specific instruction, participants select a decision threshold that maximizes reward rate. However these results are mixed, showing that many subjects are unable to do so. In two experiments, I examine a variety of factors expected to influence a participant's ability to maximize reward rate.

Feedback is widely considered an important factor in successfully learning a new skill. However in some instances learning can occur without explicit feedback, perhaps as the result of self monitoring. It is not clear the extent to which feedback plays a role in helping subjects adopt an optimal speed-accuracy tradeoff policy. While most of the literature on optimal threshold setting has been conducted with experiments that provide feedback after each trial, it is not uncommon to perform similar experiments without feedback (e.g., Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009).

This question is partially addressed by Starns and Ratcliff (2010), who examined performance on a numerosity judgement task. The authors found that indeed, decision thresholds were closer to optimal in the feedback condition compared to the no-feedback condition for younger participants, whereas older participants showed a strong preference for accuracy, and thus did not show a difference in decision threshold.

The current experiment differs in that feedback is manipulated on a within-subjects basis. This provides the opportunity to quantify the performance gain for feedback, and assess changes in performance pre- and post-feedback. If feedback serves as a mechanism for learning the task, changes in drift rate and decision threshold should be apparent after feedback is introduced.

Block length is another factor that may affect one's ability to manage time effectively. Research in vigilance has shown that performance on a repetitive task will decline over time (e.g., Mackworth, 1948), a result of fatigue and inattention. It is not known how this vigilance decrement affects cognitive parameters of the diffusion model.

Another aspect which may affect selection of a decision threshold is the *variability of trial difficulty*. In the experiments reviewed above, trial difficulty was kept constant throughout a block of trials. In this case, the optimal solution is for a participant to set a new threshold that maximizes performance for each block (difficulty level). However, when the difficulty of trials varies within a block, a new strategy is possible: participants may choose to respond more quickly on difficult trials in hopes of getting an easier trial. That is, there are additional tradeoffs between trial types that have not been explored in the extant research. A generalization of the reward rate maximization theory used in optimal foraging theory (Equation 5) allows for the possibility of maximizing performance across a block of trials with multiple difficulties. The implication here is that people may perform differently simply by shuffling the order of trials, behaving one way when trials are grouped into easy and difficult blocks, and another way when trials are intermixed. This may have important consequences for experimental design.

Some evidence suggests that participants do indeed perform differently when trials of varying difficulty are mixed within a single block. In a word naming task, response times are slower and error rates are lower for easy trials when mixed with hard trials; conversely response times are faster and error rates are higher for hard trials when mixed with easy trials (Rastle, Kinoshita, Lupker, & Coltheart, 2003; Lupker, Brown, & Colombo, 1997). Unlike Balci et al. (2011), this result suggests that participants can make adjustments in decision threshold between blocks. However it is still unclear whether participants can modulate their decision threshold within a block, or whether a single threshold is chosen for both difficulty levels.

Lastly, *foreknowledge of task difficulty* may influence a participant's speed-accuracy tradeoff. In a typical psychophysics experiment, trials are not labeled to indicate difficulty. Labeling trials as *easy* or *hard* obviates the need to estimate trial difficulty, a factor that affects the optimal decision threshold. Thus although Balci et al. (2011) found that participants chose a threshold corresponding to the average difficulty level, labeling trials might assist participants in determining an appropriate speed-accuracy tradeoff policy without needing to make a subjective assessment. Foreknowledge of task difficulty may be particularly important in conjunction with variability of trial difficulty. It seems plausible that subjects may adopt separate thresholds for different trial difficulties only when they can be sure these differences are legitimate, and not due to noise.

Pilot Study

A pilot experiment was conducted to assess the validity of these ideas. Ten participants completed a random dot motion experiment very similar to that described in Experiment 1 below. The key manipulations were *variability of trial difficulty* and *block length*.

The experiment consisted of two five-minute and five one-minute blocks in which participants were instructed to get as many points in the time allotted, gaining one point for a correct response and losing one point for an incorrect response. Each trial was randomly assigned a difficulty level (easy, medium, or hard) which was manipulated by the coherence level of the dots (20%, 15%, and 10% respectively). In an effort to prevent participants from adjusting their speed-accuracy tradeoff throughout the experiment, feedback was not provided after each trial. However, after each block participants were informed of how many points they earned during that block. The order of blocks was randomized.

Analysis hinted at a difference in speed between the one-minute ($M = 1.8s$) and five-minute ($M = 2.11s$) blocks, $F(1, 9) = 4.21$, $p = .07$, $Cohen's f = .68$, as well as a difference in accuracy between the one-minute ($M = .85$) and five-minute ($M = .8$) blocks, $F(1, 9) = 3.61$, $p = .09$, $Cohen's f = 0.63$. Additionally, as expected, there was a linear effect of difficulty on response time ($M_{easy} = 1.77s$, $M_{medium} = 1.93s$, $M_{hard} = 2.17s$, $F(1, 9) = 13.68$, $p = .005$) and a linear effect of difficulty on accuracy ($M_{easy} = .88$,

$M_{medium} = .84$, $M_{hard} = .76$, $F(1, 9) = 56.5, p < .001$).

A diffusion model analysis was conducted in order to discern which parameters of the RDM could explain these behavioral differences. Both drift rate and decision threshold were allowed to vary by block length and trial difficulty. Note that a rise in accuracy and a decrease in response time cannot be explained by a change in decision threshold alone. The simplified RDM was used, and as such between trial variability in drift rate, variability in non-decision time, and variability in starting point were not included in the model. Starting point was constrained to be halfway between the upper and lower decision thresholds, and non-decision time was estimated once for each subject and did not vary across blocks or trial difficulty.

An analysis of decision thresholds revealed an interaction between block length and trial difficulty, $F(2, 18) = 3.74$, $p = .044$, depicted in Figure 5. At face value, it appears that participants do not alter their decision threshold by trial difficulty in the five-minute blocks, but do in the one-minute blocks. However, it is not immediately clear why this would be the case. A main effect of block length on decision threshold was not observed, $F(1, 9) = 0.02$, $p = .9$, nor was a main effect of trial difficulty on decision threshold, $F(2, 18) = .86$, $p = .44$.

As expected, the diffusion model analysis uncovered a linear relationship between trial difficulty and drift rate, with harder trials having lower drift rates, $M_{easy} = .137$, $M_{medium} = .09$, $M_{hard} = .052$, $F(1, 9) = 7.57$, $p = .02$. Analysis hinted at lower drift rates in the five-minute block compared to the one-minute block, $M_{1min} = .104$, $M_{5min} = .082$, $F(1, 9) = 3.53$, $p = .09$, $Cohen'sf = .66$. This result indicates that a difference in performance between blocks may not be due to a change in the speed-accuracy tradeoff policy of the subject, but that trials in the five-minute block are in some sense harder than corresponding trials in the one-minute block, despite using identical stimuli. A plausible explanation for this effect is fatigue, which leads to a decrease in vigilance (i.e., inattention) during longer blocks. This is a novel effect not previously modeled with the

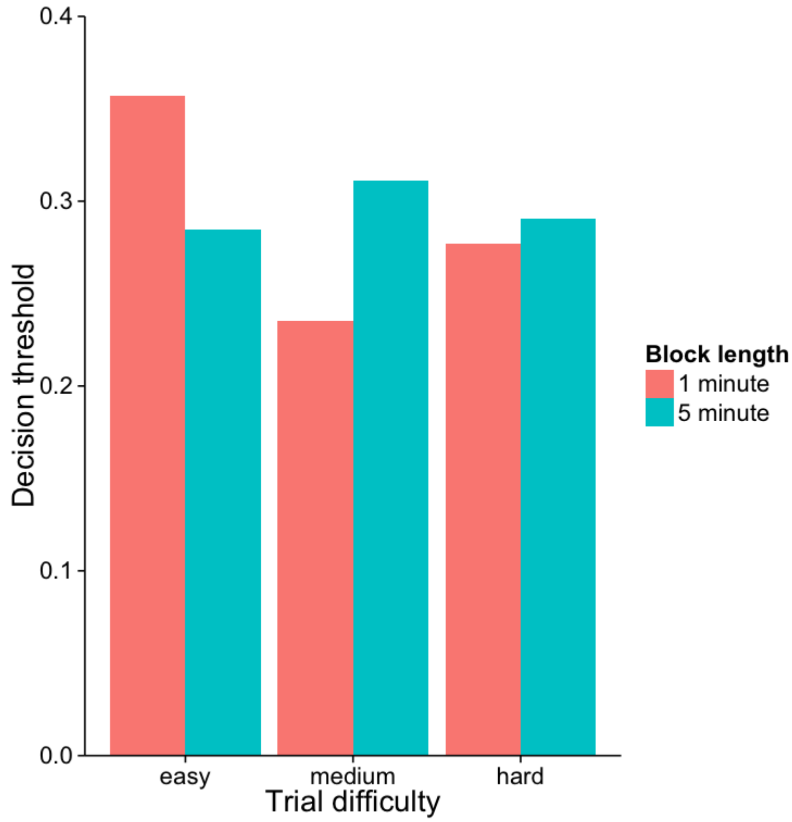

Figure 5. Analysis of decision threshold revealed an interaction between trial difficulty and block length.

diffusion paradigm. This result may have implications for experimental design. Block lengths for a typical psychology experiment are often arbitrary and usually do not vary throughout an experiment. Though experimenters frequently choose reasonable block lengths and incorporate breaks, this result suggests that block length may have a discernible effect on performance even at relatively short durations.

Another question of interest was whether participants selected an appropriate decision threshold, in accordance with the reward rate maximization theory. For this analysis, a separate diffusion model was fit to the data in the same manner as above, but collapsing across block length. Thus for each participant, three drift rates and three decision thresholds were computed, corresponding to the three difficulty levels in the task. Optimal thresholds were computed using the procedure in Equation 5, which yields a different optimal threshold for each difficulty level. Because incorrect responses were explicitly penalized, the gain function (Equation 6) was modified to account for this:

$$G^* = G(v, a) - (1 - G(v, a)) \tag{8}$$

Note that this formulation is very similar to the modified reward rate theory tested by Bogacz et al. (2010) in which errors are implicitly penalized. Unlike that experiment, errors in the current experiment *are* penalized explicitly. Thus, the theory becomes a normative one rather than descriptive one. Additionally, Bogacz et al. (2010) use a free parameter to weight penalties, whereas the above formulation contains no weighting (i.e., no free parameter).

Results are shown in Figure 6 (red circles). It appears that some of the decision thresholds are close to optimal, but many are not ($RMSE = .72$, though $RMSE = .1$ for five subjects closest to optimal[1]). In particular, many subjects chose a conservative decision threshold, high above the optimal value. This strategy results in a higher overall accuracy, but does not maximize the total number of points gained. This result is somewhat counter-intuitive, as participants did not receive any feedback with regards to accuracy on any individual trial, but were informed of the total number of points earned. Thus one might expect subjects to be more sensitive to points accumulated rather than overall accuracy. While previous findings suggested that participants do err on the side of accuracy when choosing a decision threshold, these results are more extreme than those seen previously. One notable difference is that most previous experiments incorporate feedback (e.g., Bogacz et al., 2010), while the current experiment did not. Feedback may play a crucial role in allowing subjects to determine their optimal speed-accuracy tradeoff policy. It may be that subjects simply act conservatively when they are uncertain of their own

[1]RMSE was calculated by first collapsing each participant into a single data point, the average of three decision thresholds. This was done to be consistent with the following analysis and to meet assumptions of independence between data points.

performance. This notion is supported by Balci et al. (2011), where decision thresholds were initially too high, but approached optimal thresholds after prolonged experience with the task. It is worth noting that despite the lack of feedback, most subjects improved greatly from after an initial practice session, suggesting that subjects are still able to learn somewhat in the absence of feedback. However a systematic comparison between feedback and no-feedback conditions can show the extent to which feedback plays a role in helping participants adopt an optimal strategy.

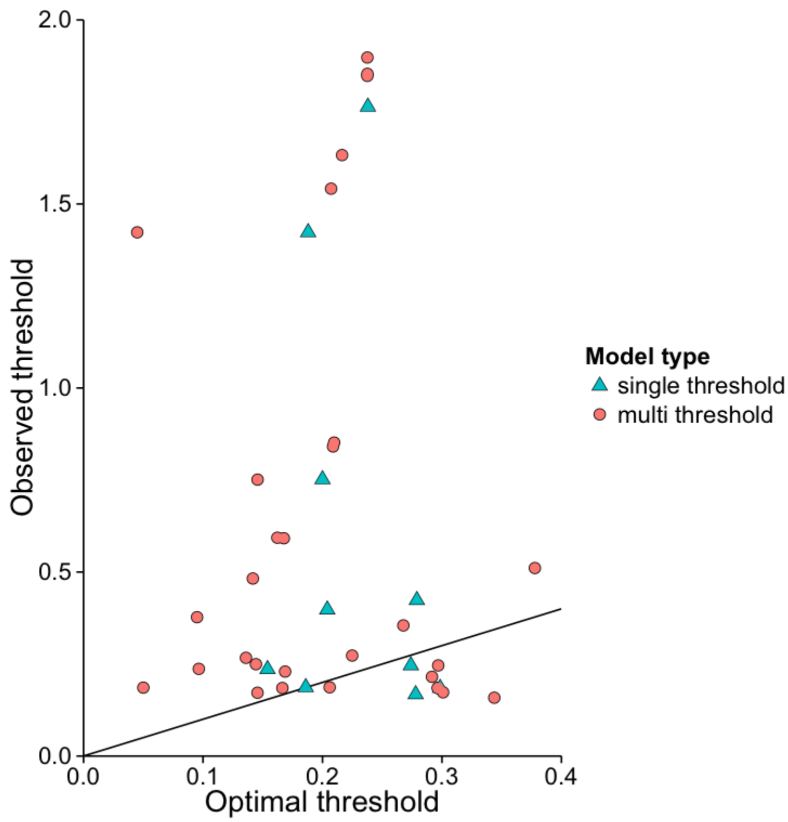

Figure 6. The identity line (solid black) shows perfect correspondence to an optimal speed-accuracy tradeoff policy. Two theories are compared: red circles indicate a single threshold for all difficulty levels, whereas green triangles indicate a separate threshold for each difficulty level. While many of the data points cluster around the identity line, it is clear that others are far from optimal.

Alternative “optimal” thresholds were also computed using Equation 2, under the assumption that subjects might choose a single threshold for all trials, regardless of difficulty. Thus for each subject, a single decision threshold was estimated rather than three. The results are shown in Figure 6 (green triangles). When thresholds were computed in this manner, results were similar to those previously computed ($RMSE = .651$, though $RMSE = .08$ for five subjects closest to optimal). Overall, it is not clear whether there is a discernible difference between the two methods of computing decision thresholds, and thus it is not clear which strategy participants are adopting. The two strategies make similar predictions, thus more data may need to be collected in order to discriminate between them.

Experiment 1

Overview

The pilot experiment found that a block length manipulation led to a vigilance decrement in performance. Furthermore, a diffusion model analysis suggested this decrement could be attributed to a decrease in drift rate for longer blocks, perhaps as a result of decreasing attentional resources. Quantifying vigilance decrements as a change in drift rate suggests a novel way of interpreting and tracking changes in attention over time. However these effects did not reach statistical significance in the pilot study due to the small sample size (N=10). The current study is designed to replicate this effect and further explore the connection between the cognitive parameters of the diffusion model and behavioral changes resulting from a vigilance manipulation.

Furthermore, the pilot study suggested that participants set their decision threshold conservatively high in the absence of feedback. Limited research has explored the connection between feedback and diffusion model parameters, despite that feedback is commonly manipulated in psychological experiments. The current study extends the work of Starns and Ratcliff (2010) by using a within-subject feedback manipulation to control for individual differences. Additionally, the within-subject manipulation allows an inspection of whether feedback simply engages attention (a temporary boost in performance only when feedback is present), aids in learning (a persistent boost in performance even after feedback is removed), or both.

Subjects

Forty Rice University undergraduate students (22 female, 18 male) were recruited to participate in the experiment in exchange for credit towards a course requirement. All subjects were required to have normal or corrected-to-normal vision. As incentive, three participants with the best performance were rewarded with $25 each at the completion of data collection.

Stimuli

Subjects participated in a random dot motion experiment (Britten et al., 1992) using stimuli generated from the Python psychophysics library VisionEgg (Straw, 2008). On each trial of the experiment, participants were shown a 300x300 square of 200 moving dots. Each dot measured 3x3 pixels, had a velocity of 100 pixels per second, and a lifespan of 80ms (i.e., each dot disappeared after 80ms and reappeared in a random location). A small proportion of dots (known as the coherence level) always moved in a pre-determined direction (left or right), while the remaining dots move in a random direction; a higher coherence level makes the task less difficult. The task of the participant is to determine the net moving direction of the dots (left or right), and respond accordingly. For the current experiment, the coherence level was set to .2, indicating that 20% of the dots moved coherently either to the right or left. Trial difficulty did not vary throughout the experiment. An inter-stimulus interval of 1s was used between trials, during which a fixation cross was presented.

Design

This experiment manipulated *feedback* and *block length* on a within-subjects basis. Some blocks gave participants feedback after each trial to inform them if the trial was answered correctly, while other blocks contained no feedback. Similarly, some blocks allowed participants five minutes to correctly answer as many trials as possible, whereas other blocks allowed participants only one minute.

Feedback order was manipulated between subjects. While all subjects completed both feedback and no-feedback blocks, half of the subjects received feedback blocks first whereas the other half of the subjects completed the no-feedback blocks first.

Behavioral analysis looked at the effect of these manipulations on response time and accuracy. Additional modeling analysis examined the effect of these manipulations on diffusion parameters (non-decision time, drift rate, and decision threshold) in order to interpret any behavioral changes. Lastly, optimal thresholds for each condition were computed and compared to observed decision thresholds.

Procedure

Participants completed ten minutes worth of practice trials (without feedback) to become acquainted with the task. Pilot studies indicated that performance improves rapidly in the first few minutes of the task; this practice block was used to minimize any practice effects.

The remainder of the experiment was divided into two halves. In one half, subjects received feedback after each trial to indicate whether they answered the trial correctly or incorrectly. A tone was played after an incorrect answer, but otherwise an incorrect response was not penalized either explicitly (loss of points) or implicitly (a time delay). These blocks were counterbalanced across subjects, such that half of the subjects received the feedback block first, and half received the no-feedback block first.

In each block, subjects completed two five-minute sub-blocks and ten one-minute sub-blocks. These sub-blocks were pseudo-randomized to minimize order effects, such that each quarter of the experiment contained one five-minute block and five one-minute blocks.

Participants were instructed to get as many trials correct in the time allotted, and told that in order to achieve this they must respond both quickly and accurately. These instructions were emphasized to mitigate any implicit bias the subjects may have with regard to experimenter demand—i.e., that speed is more important than accuracy or vice versa.

After each sub-block, participants were told how many points (correctly answered trials) they earned for that round. This was done after every sub-block in the experiment, and was not considered part of the "feedback" manipulation. In between sub-blocks, participants were allowed a short break if they desired, or they were able to continue with the experiment immediately.

Afterwards, participants completed a short demographic questionnaire containing a few questions related to their beliefs about their strategy and performance in the task. Specifically, participants were asked 1) if they felt the block length affected their response time, and 2) if they felt the presence of feedback affected their accuracy. At the completion of the one-hour experiment, participants were told their cumulative score for the experiment.

Outlier Removal

The data were filtered for outliers prior to data analysis. Following Donkin, Averell, Brown, and Heathcote (2009), all trials with a response time of less than 180ms or greater than 10s were removed from the dataset. These trials were excluded as they are not often adequately modeled by the diffusion process. Trials faster than 180ms may indicate eager motor responses (guesses), whereas trials slower than 10s may indicate prolonged inattention.

Additionally, six subjects whose accuracy was below 60% (after outlier removal) were not included in the analyses. Most participants with accuracy below 60% also had a high proportion of outlier RTs, and nearly all of the outliers from these subjects were removed for being fast outliers (0 to 180ms). Of the remaining 34 participants, 1.4% of the trials were removed as outlier RTs. Fortuitously, half of the remaining subjects received feedback first, whereas half received feedback second.

Model Fitting

Parameters of the drift diffusion model were fit separately for each participant and each condition using DMAT (Diffusion Modeling Analysis Toolbox; Vandekerckhove & Tuerlinckx, 2008, 2007) for MATLAB. Non-decision time, drift rate, and decision threshold were allowed to vary across all conditions. Starting point was fixed between the two decision boundaries, which assumes that participants are equally likely to choose rightward or leftward motion. This decision is justified in that the dots were equally likely to move rightward or leftward, and responses were rewarded the same regardless of direction of motion.

The simplified version of the Ratcliff diffusion model was used, meaning between-trial variability in the parameters (i.e., s_z, s_v, and s_{ter} in Table 1) was not included in the model. The primary purpose of this was to be able to calculate and compare optimal thresholds to observed thresholds. When these parameters are included in the model, an analytic solution to the calculation of optimal thresholds is no longer possible. Overall, the set of parameters used was similar to that of the pilot experiment with the exception that non-decision time was now allowed to vary across blocks. Though the current study's manipulations were not expected to affect non-decision time, some experimental manipulations have been shown to affect this parameter (Dambacher & Hübner, 2014; Mulder et al., 2013).

A variety of fitting procedures can be used to derive model parameters from experimental data (Heathcote, Brown, & Mewhort, 2002; Voss & Voss, 2008; Wagenmakers, van der Maas, Dolan, & Grasman, 2008; Vandekerckhove & Tuerlinckx, 2007). A comprehensive guide for parameter fitting is available from Ratcliff and Tuerlinckx (2002), which compares the robustness of several popular parameter fitting methods. The process used in the current study is described briefly:

The EZ-diffusion model (Wagenmakers et al., 2007) provides a closed-form analytic solution to fitting a three-parameter diffusion model (drift rate, non-decision time, and decision threshold) using only error rate, mean response time, and variance in response time. While the EZ-diffusion fitting procedure has known limitations (Ratcliff, 2008; Wagenmakers et al., 2008), it is used to provide initial parameter settings for a more robust fitting procedure.

Using these parameters, a theoretical cumulative distribution function is generated; that is, a function that relates decision time to the probability that the diffusion process has terminated (see Ratcliff & Tuerlinckx, 2002, Appendix A). This theoretical distribution can be compared to the observed distribution using a cost function. Historically, the most common approach is to divide the distribution into quantiles and compute a chi-square statistic. Though this approach was favored because it is less computationally demanding than other approaches, maximum likelihood estimation has become a popular alternative in recent years (Vandekerckhove & Tuerlinckx, 2007). Though each cost function has certain advantages and disadvantages, the maximum likelihood method was chosen for the current set of experiments as it is often adequate and under certain conditions can be the most accurate method for recovering diffusion parameters (Ratcliff & Tuerlinckx, 2002).

The parameters of the model are then adjusted by small increments using the SIMPLEX algorithm (Nelder & Mead, 1965) in order to find the parameters that maximize the likelihood of the data. The search is terminated after a fixed number of iterations or when a given tolerance is reached. For the current experiments, the default procedure provided by DMAT was used. After the model parameters for each subject and condition are derived, differences in parameters between conditions can be assessed using traditional statistical methods (e.g., repeated-measures ANOVA).

Behavioral Results

Response Time. A linear mixed-effect model was used to assess the effects of block length, feedback, and feedback order on response time. Subject and all subject by within-subject factors (feedback and block length) were treated as random effects. This was achieved using the *lme4* and *lmerTest* packages (Bates, Maechler, Bolker, & Walker, 2014; Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2014) in the R programming language (R Core Team, 2014). Linear mixed-effect models can offer more power than a traditional repeated-measures ANOVA analysis, though their use in psychological fields is nascent. For clarity, the R syntax used here and throughout is provided in Appendix A.[2]

Overall mean response time did not differ between subjects who received feedback first and those who received feedback second ($p = .39$). As seen in Figure 7, participants who received feedback in the second half of the experiment did show a decrease in response time when feedback was introduced ($M_{fb} = .92s$, $M_{nofb} = 1.23s$), while those who received feedback in the first half of the experiment had similar response times when feedback was removed ($M_{fb} = 1.08s$, $M_{nofb} = 1.07s$). This feedback by feedback order interaction was significant ($\beta = .35$, $SE = .17$, $t = 2.08$, $p = .046$). Only 47% of subjects who received feedback first showed a decrease in mean response time with feedback, whereas 88% of subjects who received feedback second showed a decrease in mean response time.

It is perhaps surprising that feedback produced shorter response times for those participants who received feedback in the second half of the experiment rather than the first half. This could be explained as a confluence of feedback and practice effects (Dutilh et al., 2009; Starns & Ratcliff, 2010). Collapsing across conditions, mean response time decreased steadily across four quarters of the experiment ($M_{q1} = 1.23s, M_{q2} = 1.08s$, $M_{q3} = 1.02s, M_{q4} = 0.96s$). Thus, it seems plausible that for those who received feedback

[2] A separate analysis was conducted on log transformed response times, to determine if the skewed distribution of response times significantly affected the results. All results were qualitatively identical; that is, all significant effects remained significant and all non-significant effects remained non-significant.

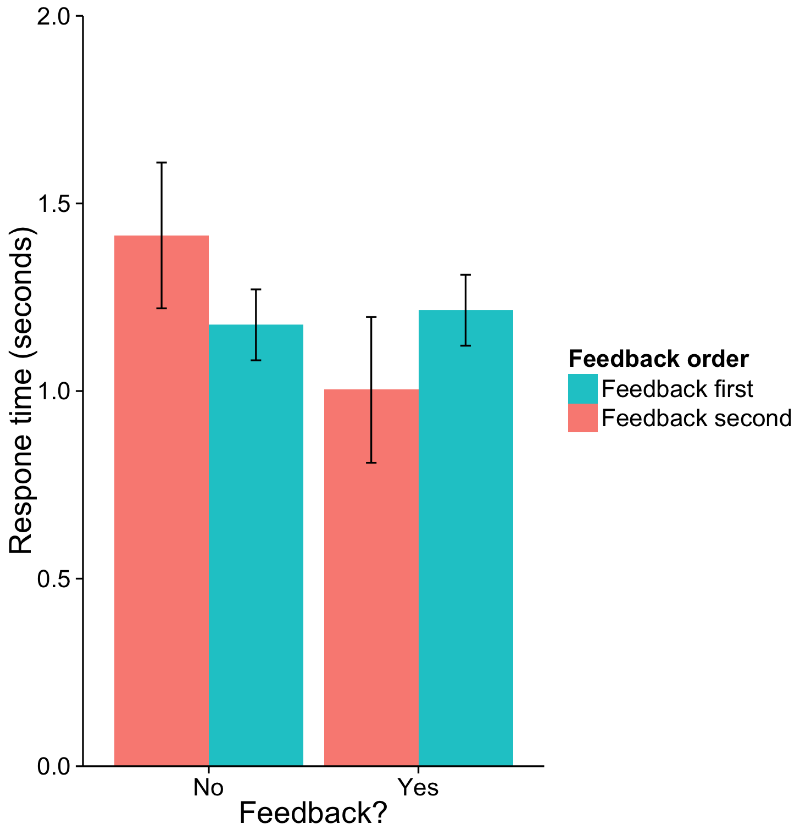

Figure 7. Participants who received feedback first showed little change in response time when feedback was removed, whereas participants who received feedback second showed a dramatic decrease in response time when feedback was introduced. Though response times decreased over time for both groups, the group with feedback had shorter response times in both the first half (outer bars) and second half (inner bars) of the experiment. Error bars here and throughout indicate 95% within-subject confidence intervals (Morey, 2008).

second, practice effects and feedback worked in conjunction with each other; for those who received feedback first, the two effects are opposed.

More generally, providing feedback after incorrect responses led to a decrease in response time ($M_{fb} = 0.99s$, $M_{nofb} = 1.15s$, $\beta = -.34$, $SE = .12$, $t = -2.91$, $p = .007$). Additionally, one-minute blocks produced shorter response times than five-minute blocks ($M_{1min} = 1.01s$, $M_{5min} = 1.13s$, $\beta = .28$, $SE = .048$, $t = 5.82$, $p < .001$). 67% of subjects showed a decrease in mean response time with feedback, whereas 91% of subjects showed a

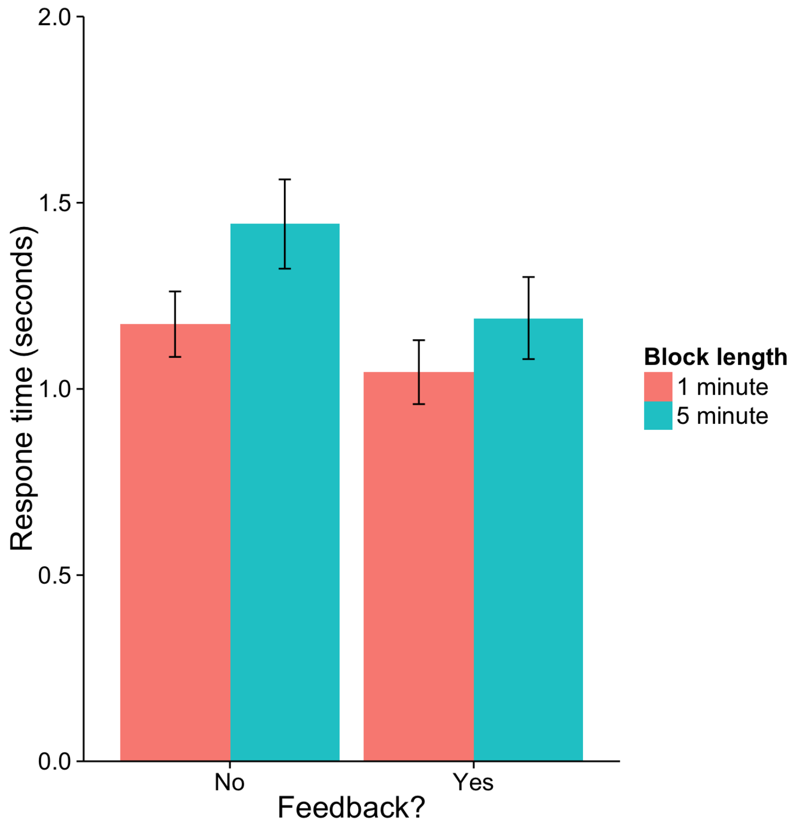

Figure 8. Response times were shorter both with feedback and in shorter block lengths. However feedback appears to have a larger effect in the five-minute block compared to the one-minute block.

decrease in mean response time with a shorter block length.

Feedback resulted in a larger decrease in response time for the five-minute block compared to the one-minute block (difference in means: $M_{5min} = 0.21s$, $M_{1min} = 0.1s$; see Figure 8). This feedback by block length interaction was significant ($\beta = -.22$, $SE = .024$, $t = 9.03$, $p < .001$). This suggests that the effects of feedback and block length are not linearly additive, perhaps because there is a lower bound on response time.

Lastly, a three-way interaction of feedback, block length, and feedback order was significant ($\beta = .23$, $SE = .034$, $t = 6.81$, $p < .001$), seen in Figure 9. It seems participants who begin the experiment with feedback show no change in response time when feedback is removed, but still do show an effect of block length. Those participants who start the experiment without feedback begin with the slowest response times, but after introducing feedback these same participants have the fastest response times. An effect of block length appears present everywhere except perhaps during feedback blocks for participants who receive feedback second (far right in Figure 9).

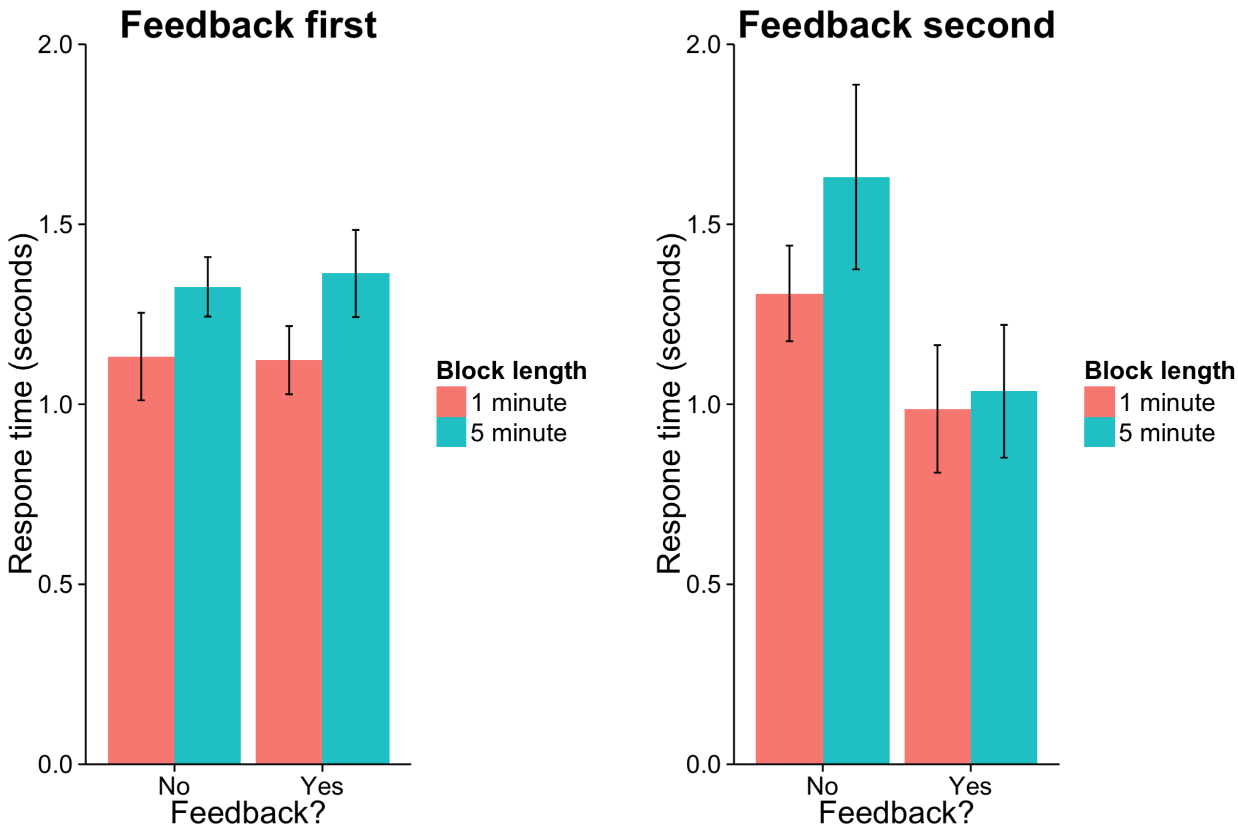

Figure 9. A significant three-way interaction of feedback, feedback order, and block length on response time was detected.

Accuracy. A corresponding linear mixed-effect model was used to assess the effects of block length, feedback, and feedback order on accuracy. Subject and all subject by within-subject factors (feedback and block length) were treated as error terms.

A feedback by feedback order interaction, shown in Figure 10, was nearly significant ($\beta = -.039$, $SE = .021$, $t = -1.87$, $p = .069$). 41% of participants who received feedback first had a higher mean accuracy in the feedback blocks, whereas 65% of participants who received feedback second showed an improvement in accuracy when feedback was

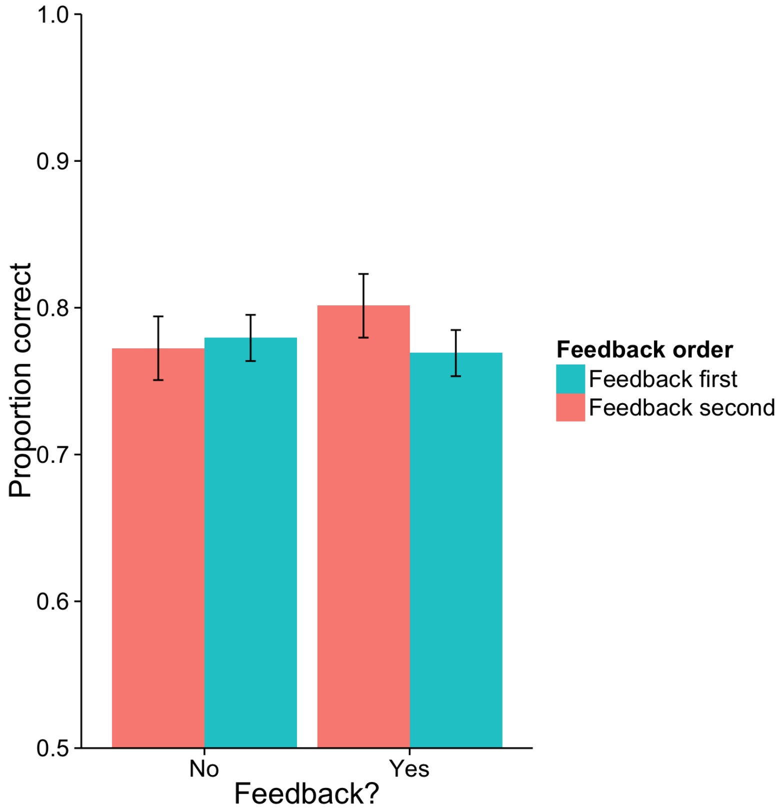

Figure 10. Though accuracies improved over time for both groups, the group with feedback had a higher mean accuracy in both the first half (outer bars) and second half (inner bars) of the experiment

introduced. Analogous to the response time data, these results could be partially attributed to practice effects in which feedback and practice work in conjunction for those who receive feedback second, and are opposed for those who receive feedback first. Overall, accuracy improved slightly with feedback ($M_{fb} = .829$, $M_{nofb} = .816$), however this effect was not quite significant ($\beta = .028$, $SE = .015$, $t = 1.92$, $p = .061$). Only 53% of subjects showed an improvement in accuracy with feedback.

Though the pilot experiment hinted at a main effect of block length on accuracy, the current study provided no support for this hypothesis ($p = .48$). At first glance, it may appear that this effect is masked by a significant feedback by block length interaction ($\beta = .023$, $SE = .011$, $t = 2.16$, $p = .031$). As seen in Figure 11, mean accuracy is roughly

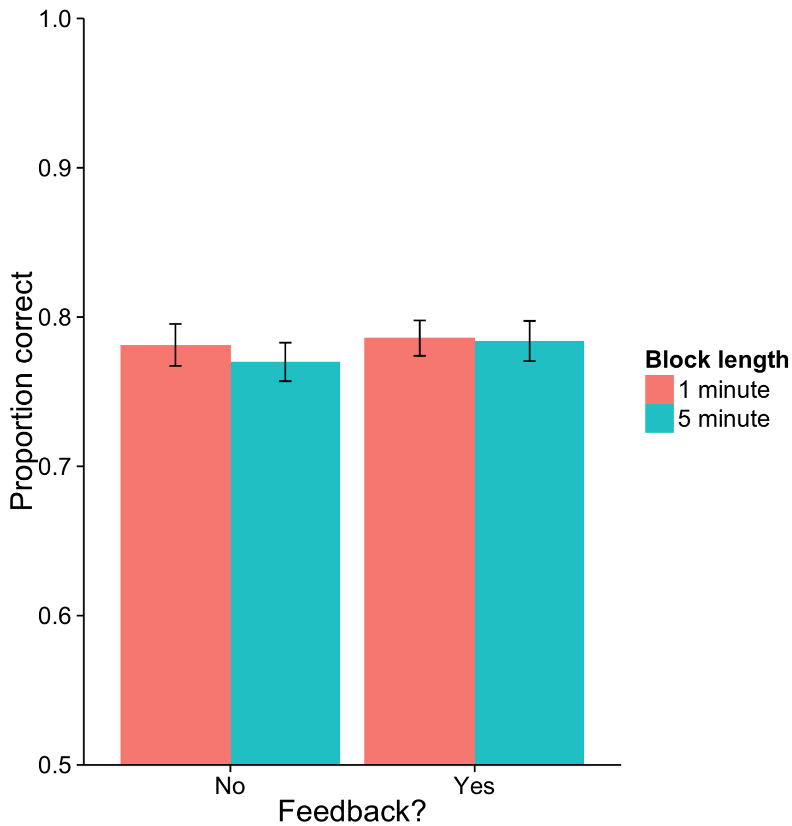

Figure 11. The difference in accuracies between the one-minute and five-minute blocks was significantly larger for no-feedback compared to feedback trials.

1% higher in one-minute compared to five-minute blocks for no-feedback trials. The pilot experiment did not provide feedback, and thus this would explain the discrepancy; however a simple effects analysis of only no-feedback trials did not support this hypothesis ($p = .61$).

Lastly, a near-significant three-way interaction between block length, feedback, and feedback order is present ($\beta = -.028$, $SE = 1.53$, $t = -1.86$, $p = .063$), seen in Figure 12. Overall, the experimental manipulations (block length, feedback, and feedback order) had similar effects on both response time and accuracy.

Modeling Results

Diffusion model parameters (non-decision time, drift rate, and decision threshold) were derived for each subject using the methodology described above. Separate parameters

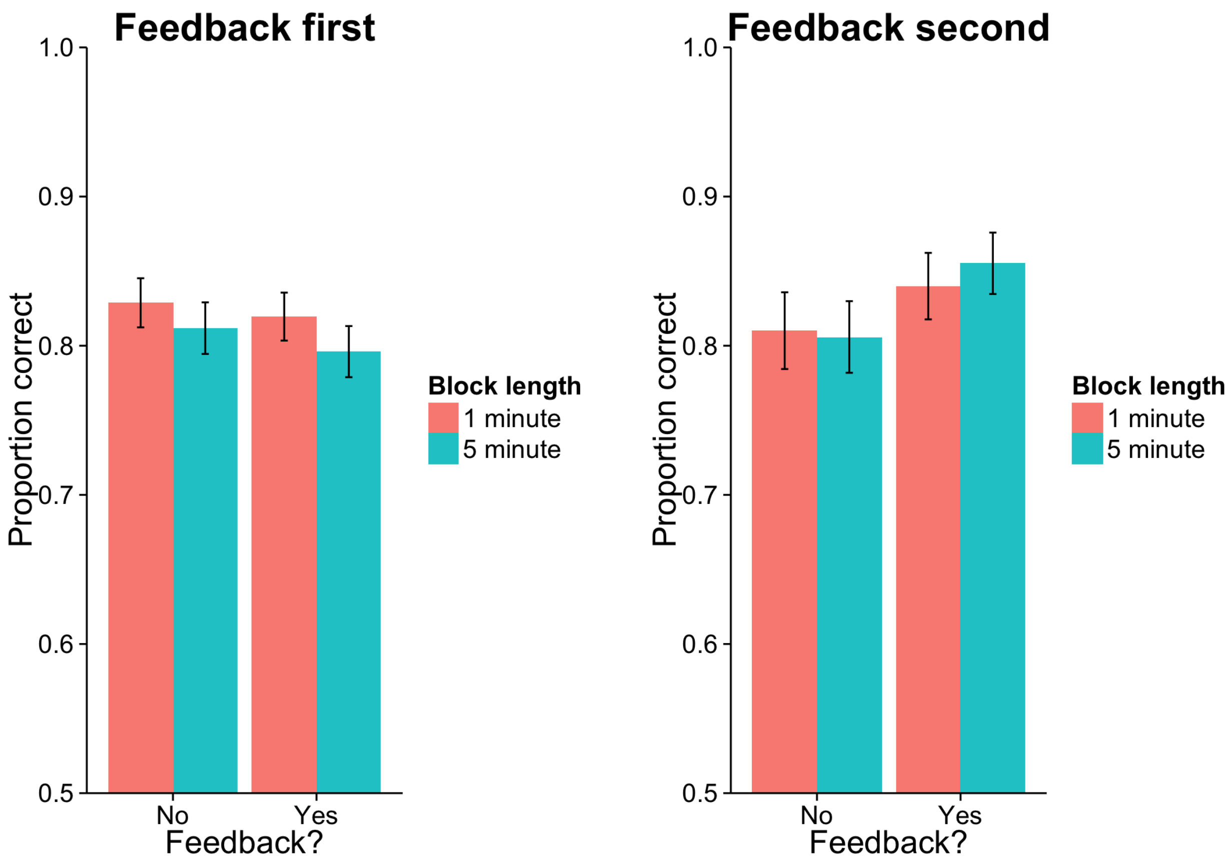

Figure 12. A three-way interaction between block length, feedback, and feedback order was detected.

were calculated for each of the four cells (feedback by block length) in the design. Model parameters were then compared between conditions using a repeated-measures ANOVA.

Non-decision Time. Non-decision time was allowed to vary across conditions, as some experiments have found instructional changes to affect this parameter (Dambacher & Hübner, 2014; Mulder et al., 2013). Though non-decision time was not expected to change with the manipulation of feedback or block length, allowing the parameter to vary and then assessing a change across conditions is perhaps a more conservative approach. As expected, however, there were no main effects or interactions that affected non-decision time.

Drift Rate. A significant interaction between feedback and feedback order was observed, $F(1, 32) = 20.32$, $MSE = .032$, $p < .001$. Drift rate increased significantly with

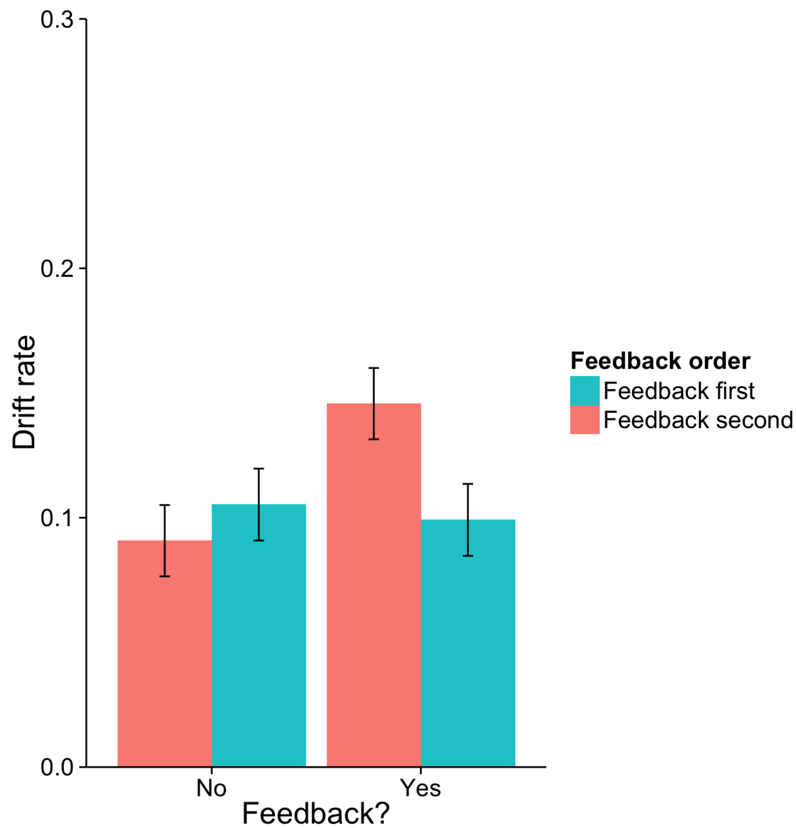

Figure 13. Drift rate increased over time for both groups, though the group with feedback had higher drift rates in both the first half (outer bars) and second half (inner bars) of the experiment.

feedback for those who received feedback second, but remained relatively constant for those who received feedback first (Figure 13). 100% of participants who received feedback second showed an increase in drift rate with feedback, whereas only 52% of those who received feedback first showed an increase in drift rate during feedback blocks. Collapsing across feedback order, feedback resulted in a slight increase in drift rate $M_{fb} = 0.122$, $M_{nofb} = 0.098$, $F(1, 32) = 12.98$, $MSE = .02$, $p = .001$. 76% of subjects showed higher drift rate during feedback blocks. This suggests that providing feedback may improve one's perceptual ability to discriminate between stimulus classes.

Drift rates appeared to be slightly lower in the five-minute compared to one-minute blocks; however, this effect was not quite significant, $M_{1min} = .115$, $M_{5min} = .106$,

$F(1, 32) = 3.87$, $MSE = .002$, $p = .058$. 65% of subjects displayed higher drift rates in the one-minute compared to five-minute blocks.

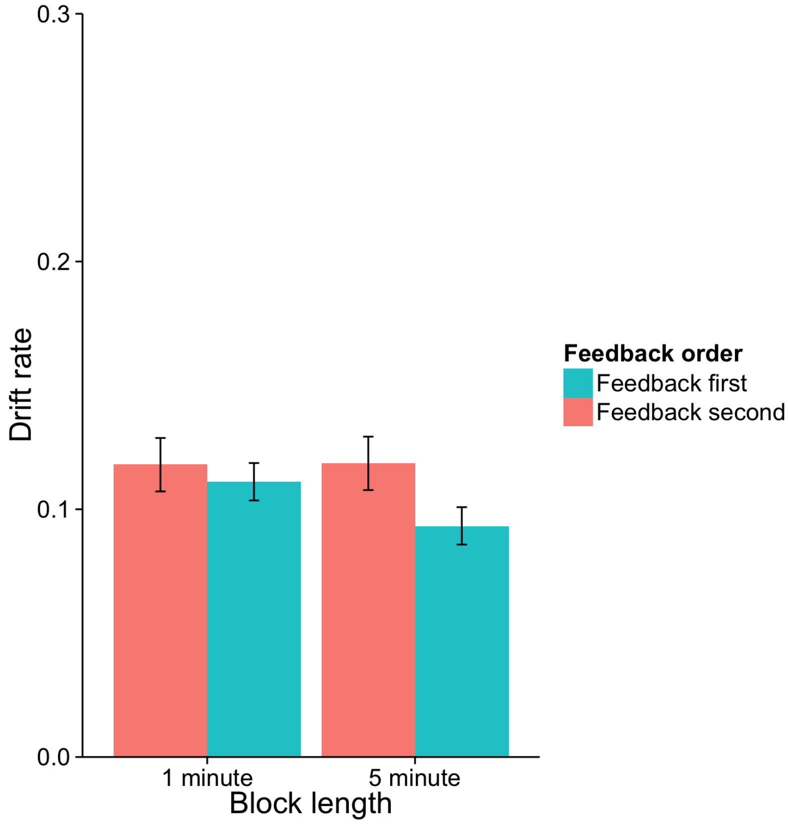

Figure 14. Drift rates decreased with longer block lengths, but only for participants who received feedback first.

Lastly, block length interacted significantly with feedback order, $F(1, 32) = 4.37$, $MSE = .003$, $p = .045$. Those who received feedback second showed similar drift rates for the one-minute and five-minute blocks, whereas those who received feedback first had lower drift rates in the five-minute compared to the one-minute blocks (Figure 14). It is not immediately clear why this would be the case.

Decision Thresholds. Overall, blocks with feedback resulted in lower decision thresholds, $M_{fb} = .187$, $M_{nofb} = .215$, $F(1, 32) = 4.7$, $MSE = .028$, $p = .038$. 62% of subjects showed lower thresholds during feedback compared to no-feedback blocks. However, this effect should be viewed in light of a significant feedback by feedback order interaction, $F(1, 32) = 4.21$, $MSE = .025$, $p = .048$ (Figure 15). It appears decision thresholds are fairly consistent with the exception of participants who have not yet received any feedback blocks. It seems likely that these participants set their decision threshold conservatively high in the absence of feedback. For those who received feedback second, 76% of participants showed lower thresholds with feedback. In contrast, only 47% of participants who received feedback first had lower thresholds during feedback blocks.

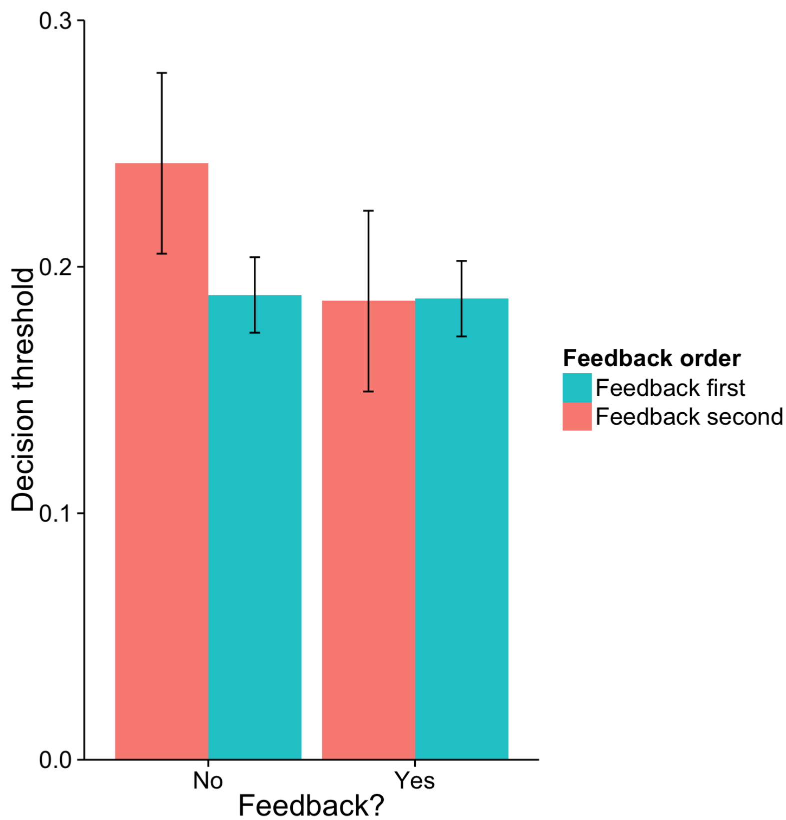

Figure 15. Decision thresholds were similar in all conditions, except for participants who had not yet done any feedback trials (far left).

Additionally, shorter block length significantly lowered decision thresholds, $M_{1min} = .188$, $M_{5min} = .214$, $F(1, 32) = 6.99$, $MSE = .022$, $p = .013$. 76% of subjects showed lower thresholds in the one-minute compared to five-minute block. This effect was anticipated, but not found in the pilot experiment. It is possible this effect was masked by a significant block length by difficulty interaction in the pilot experiment (Figure 5).

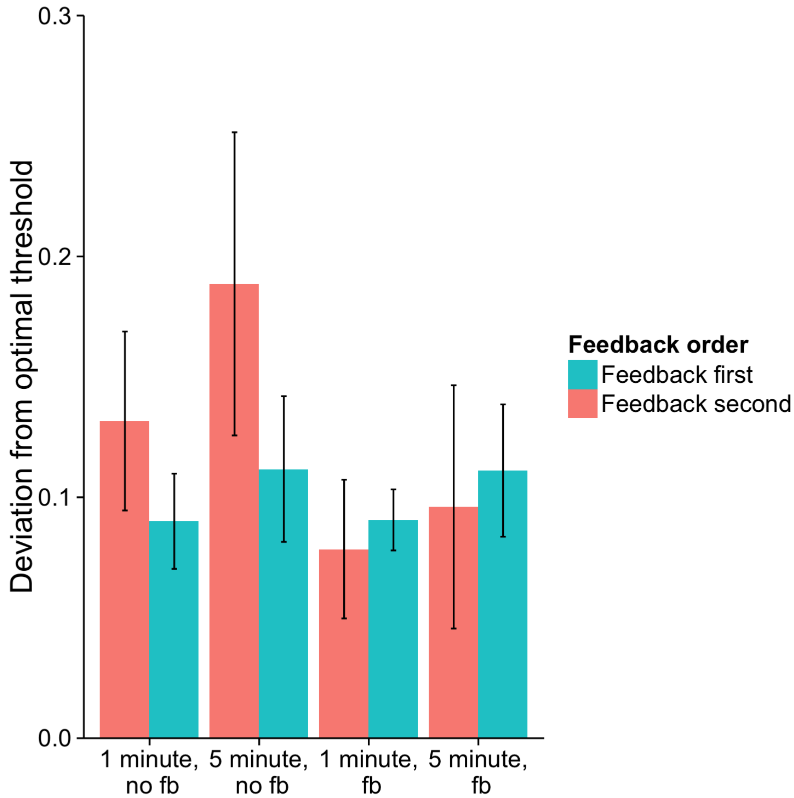

Figure 16. Decision thresholds are closest to optimal for feedback and one-minute blocks. As seen in the feedback first group, removal of feedback did not appear to worsen decision thresholds.

A more important question is how decision thresholds deviate from optimal thresholds. Optimal decision thresholds were determined for each subject and condition by finding the decision threshold that maximizes reward rate (Equation 1) given the obtained drift rate and non-decision times. This procedure is identical to the “single threshold” calculations in the pilot experiment, except that the original gain function is used (Equation 6) rather than a modified gain function (Equation 8) because incorrect answers were not penalized in this experiment. Deviation scores were then calculated, subtracting the observed threshold from the optimal threshold for each condition, and subjected to a repeated-measures ANOVA. Results are shown in Figure 16.

The effects of feedback, block length, and feedback order on optimal threshold deviations mirror the effects on decision threshold just reported. Decision thresholds are closer to optimal with feedback, $F(1, 32) = 5.61$, $MSE = .045$, $p = .024$. Again, this effect is tempered by a significant interaction between feedback and feedback order, $F(1, 32) = 5.6$, $MSE = .045$, $p = .024$. Lastly, decision thresholds are closer to optimal in one-minute compared to five-minute blocks, $F(1, 32) = 7.9$, $MSE = .029$, $p = .008$. These effects on deviation scores happen to mirror the "raw" effects on decision threshold because thresholds are systematically biased: being closer to optimal almost always entailed a lowering of decision thresholds.

Overall, decision thresholds were furthest from optimal in the no-feedback, five-minute blocks for participants who received feedback second. Thresholds were closest to optimal for the same participants in the feedback, one-minute blocks. This further supports the notion that feedback and shorter block lengths facilitate adoption of a more optimal speed-accuracy tradeoff policy.

Survey

After completing the experiment, participants were given a survey to assess their meta-knowledge of their performance. Two participants did not complete the survey. Of particular interest is whether participants were aware of the effect of block length or feedback on response time and accuracy, respectively.

First, participants were asked: "Do you feel that the block length affected the speed of your response? (i.e., did you respond faster or slower in the one-minute compared to the five-minute block?)" Ten subjects believed their response time did not differ by block length, though response time actually decreased for all ten. Twenty-two subjects believed that block length did have an effect on response time, and response time decreased for nineteen of them.

Response time analyses were re-run after dividing the data into two groups: those who believed block length had an effect on their response time, and those who did not.

However, the significance of all main effects and interactions involving block length did not differ between the two groups. In particular, a main effect of block length, an interaction of block length and feedback, and a three-way interaction of block length, feedback, and feedback order were all significant in both groups. These effects are the same as those reported initially. Thus, it seems that while a majority of participants (59%) correctly inferred that block length affected their response time, performance did not differ between participants who were or were not aware of this. Note that the wording of the question is ambiguous in that it is non-directional: participants could answer "yes" regardless of the whether they believed a shorter block length increased *or* decreased their response time. A more nuanced analysis could examine whether participants correctly inferred whether their response time differed significantly regardless of direction. Such an analysis, which might require a separate significance test for each subject, seems unlikely given the overwhelming tendency for participants to have shorter response times in the one-minute block.

Additionally, participants were asked: "Do you feel you were more accurate in the blocks that provided feedback?" Nine participants believed that their accuracy did not improve with feedback, of which accuracy improved for five (i.e., 44% were correct). Twenty-three people believed that their accuracy did improve with feedback, and eleven of them were correct (i.e., 47% were correct). Again, the data were divided into two groups based on their responses and accuracy analyses were re-run. At face value, it appears that subjects who responded "yes" correctly inferred that feedback had a positive effect on accuracy ($p = .008$) whereas feedback had no effect for those who responded "no" ($p = .37$). However in light of the tabulation above, it seems this effect is likely driven by the magnitude of the performance increase for some subjects compared to others. Overall, it seems participants are not well-calibrated when it comes to their meta-knowledge of whether feedback improves performance.

Discussion

A summary of the main effects found in the experiment are presented in Table 2. Feedback appeared to both increase drift rate and lower decision thresholds. Both of these factors explain why response times are lower for feedback blocks, whereas an increase in drift rate explains why accuracy is higher for feedback blocks.

Table 2

Summary of Main Effects from Experiment 1

IV	RT	Accuracy	Drift Rate	Decision Threshold
1-minute (c.f. 5-minute)	↓	–	↑°	↓
Feedback (c.f. no-feedback)	↓	↑°	↑	↓
Feedback first (c.f. feedback second)	–	–	–	–

Note. The direction of each main effect is indicated by an arrow.

° $.05 < p < .07$

Practice appears to increase drift rate throughout the experiment. Participants who receive feedback second start the experiment with the lowest drift rates and highest decision thresholds, resulting in the highest response times yet lowest accuracies. However, because these participants receive feedback in the second half of the experiment, working in confluence with practice effects, they end up with the highest drift rates, thus resulting in the lowest response times and highest accuracies. In contrast, practice appears to have little or no influence on decision thresholds over the one-hour time period tested. That is, participants naturally become better at discriminating between stimulus classes with practice (higher drift rate) but do not appear to lower their decision threshold significantly simply because of experience.

The feedback order manipulation demonstrates how cognitive parameters and behavior can be affected by feedback, but do not necessarily depend on the continuing presence of feedback. Decision thresholds were set conservatively high for participants who had no experience with feedback trials and thus had no internal barometer for assessing their own performance. Feedback enabled participants to select a more appropriate decision threshold, but continued feedback was not necessary to maintain this decision threshold. In a sense, the decision threshold is learned during the feedback phase. In contrast, it is not clear whether continued feedback is necessary to maintain drift rate levels, though actively providing feedback certainly increases drift rates.

The pilot experiment hinted at an increase in accuracy and decrease in response time for one-minute blocks compared to five-minute blocks, though neither effect was significant. The current study confirmed that response times decrease with shorter block lengths, but no change in accuracy was detected. As with feedback, this decrease in response time is explained by both an increase in drift rate and a decrease in decision threshold for one-minute compared to five-minute blocks.

Overall, both feedback and shorter block length had qualitatively similar effects on response time, drift rate, and decision threshold. However these effects were not displayed consistently across subjects. Table 3 summarizes the percentage of subjects who show an effect in the correct direction for all significant effects.

Table 3

Percentage of Subjects Showing a Main Effect in the Correct Direction

IV	RT	Accuracy	Drift Rate	Decision Threshold
1-minute (c.f. 5-minute)	91%	–	65%	76%
Feedback (c.f. no-feedback)	67% (47/88)	53% (41/65)	76% (52/100)	62% (47/76)

Note. Numbers in parentheses indicate percentages for subjects who received feedback first or second, respectively.

Finally, both feedback and shorter block lengths pushed decision thresholds closer to optimal levels. Subjects showed a systematic bias for caution when initially setting their decision threshold, as expected from previous studies (Holmes & Cohen, 2014). The aim of the present study differs from many previous studies, in that it is not meant to posit an alternative account of why participants choose a given threshold (for example, maximizing a weighted combination of reward rate and accuracy). Rather, the current study describes how the experimental manipulations facilitate the adoption of a more optimal speed-accuracy tradeoff policy given a true normative model. While it is clear from previous research that humans deviate from this normative policy, understanding factors that edge us closer to optimality may be important for both experimental design and for designing interventions in real-world applications.

Experiment 2

Overview

Optimal foraging theory suggests that the optimal decision threshold depends not only on the speed-accuracy tradeoff function of the current trial, but on the tradeoff function of other trials as well. In the case where difficulty does not vary across trials, such as in Experiment 1, each trial's tradeoff function is relatively identical. However when difficulty varies across trials, different trial types now have unique tradeoff functions. Thus, a separate decision threshold should be computed for each level of difficulty, and these thresholds should depend on each other. The pilot study hinted at a difference in decision threshold by trial difficulty under certain conditions, however no clear pattern emerged. The current experiment tests this theory under more constrained conditions (two difficulties instead of three) and with a larger sample size.

Moreover, optimal foraging theory suggests that we should only choose separate decision thresholds when it is possible to identify trials as belonging to a specific type, and thus identify its speed-accuracy tradeoff function (Stephens & Krebs, 1987). This raises the possibility of another manipulation: subjects may be more likely to adopt separate decision thresholds for different trial difficulties when they are made explicitly aware of a trial's difficulty.

Thus, the current experiment examines the differences in performance that result from varying trial difficulty within a block, in addition to varying a subject's knowledge of trial difficulty.

Subjects

Twenty-nine subjects (18 female, 11 male) were recruited in the same manner as Experiment 1. The experiment consisted of three one-hour sessions over the course of a week. Participants were incentivized by offering $25 to the top performer at each session. This was done to motivate participants each session, even if they felt their performance was sub-par on previous sessions.

Stimuli

Stimuli were generated using the same method as Experiment 1. Two difficulty levels were used: easy trials (0.25 coherence) and hard trials (0.2 coherence). Additionally, the inter-stimulus interval was extended to 1.5s (from 1s) in order to allow participants time to assess and react to the pre-cue before a trial began. The fixation cross presented during the inter-stimulus interval was colored either green (easy trial), red (hard trial), or white (easy or hard trial).

Design

This experiment manipulated variability in trial difficulty (*block type*) and foreknowledge of task difficulty (*pre-cue*) on a within-subjects basis. Static blocks consisted entirely of a single difficulty (hard or easy), whereas mixed blocks contained trials of varying difficulties (hard and easy). Additionally, in some blocks trial difficulty was labeled using a pre-cue (by coloring the fixation cross), whereas in other blocks participants were not informed of trial difficulty.

Analysis looked at the effect of these manipulations on response time and accuracy. Additional analysis looked at the effects on diffusion parameters (non-decision time, drift rate, and decision threshold) to explain these behavioral changes. Lastly, optimal thresholds were computed in order to assess how the experimental manipulations affected participant's speed-accuracy tradeoff policy during mixed blocks.

Procedure

The procedure was similar to that of Experiment 1. In each experimental session, subjects completed a ten minute practice block (mixed, pre-cue) to become acquainted with the task. The remainder of the experiment was divided into eight five-minute blocks.

Half of these blocks used a pre-cue to label trial difficulty by coloring a pre-trial fixation cross green or red. The other half of the blocks used a white fixation cross, so that participants were unaware of the trial difficulty (no pre-cue). Additionally, block type was manipulated such that half of the blocks contained both easy and hard trials interspersed (mixed block), whereas half contained only a single difficulty level (static block). The eight blocks were pseudo-randomized such that participants received all four possible combinations (block type by pre-cue) in each experimental half.

Participants received feedback after an incorrect trial in the form of tone. After each block, participants were informed of their score for that block.

This experiment was conducted over three one-hour sessions. At the completion of each session, participants were told their score for that day. All three sessions were scheduled to occur within a timespan of approximately one week.

Outlier Removal

The data were filtered for outliers using the same procedure as Experiment 1. Three subjects whose total accuracy was below 60% were removed from the dataset entirely. Of the remaining 26 subjects, only 0.2% of trials were removed for being outlier response times.

Model Fitting

Diffusion model parameters were derived in a manner identical to that of Experiment 1. Non-decision time, drift rate, and decision threshold were free to vary across all possible conditions, as well as between sessions.

Behavioral Results

Difficulty and Session. Two separate linear mixed-effect models were fit using response time and accuracy as the dependent variables. Subject and all subject by within-subject factors (coherence, block type, session, and pre-cue) were treated as random effects.[3]

A limitation of linear mixed-effect models is that there is no omnibus test for factors with more than two levels. Session, being a ternary variable, was broken down into two factors: a contrast representing session 1 vs. session 2 (henceforth *session 2*), and a contrast representing session 1 vs. session 3 (henceforth *session 3*).

As expected, there was a main effect of difficulty level on both response time ($M_{easy} = 0.80s$, $M_{hard} = 0.88s$, $\beta = -.193$, $SE = .024$, $t = 8.11$, $p < .001$) and accuracy ($M_{easy} = .889$, $M_{hard} = .861$, $\beta = .02$, $SE = .008$, $t = 2.45$, $p = .016$). 100% of participants were more accurate in easy trials and 88% of participants had shorter response times for easier trials. Additionally, response times appeared to decrease steadily by session ($M_{session1} = 1.0s$, $M_{session2} = 0.83s$, $M_{session3} = 0.71s$). Main effects of session 2 ($\beta = -.227$, $SE = .065$, $t = 3.5$, $p = .001$) and session 3 ($\beta = -.347$, $SE = .065$, $t = 5.53$, $p < .001$) were significant. Response time was also moderated by a significant interaction of difficulty level and session 2 ($\beta = .11$, $SE = .024$, $t = 4.59$, $p < .001$) and session 3 ($\beta = .139$, $SE = .024$, $t = 5.87$, $p < .001$). See Figure 17.

[3] As with Experiment 1, an analysis was also performed on log transformed response time. Results were very similar, and differences are noted throughout.

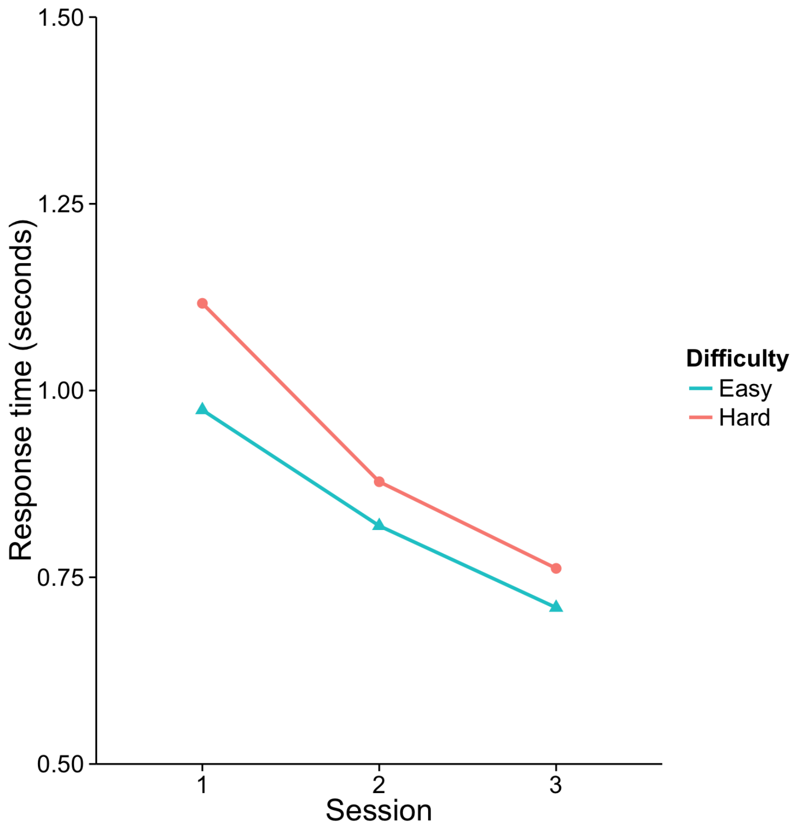

Figure 17. Response times were always higher for hard trials, though the difference in response times for hard and easy trials is largest in Session 1. Note that in this graph and all subsequent response time graphs, the y-axis is zoomed to illustrate small effects.

Block type. The main research question in this experiment it whether participants treat trials differently depending on whether they are in a block of the same difficulty (static block) or varying difficulties (mixed block).

As shown in Figure 18 and Figure 19, there were numerous interactions involving block type, including block type by session 3 on accuracy ($\beta = .031$, $SE = .011$, $t = 2.75$, $p = .006$) and response time ($\beta = -.088$, $SE = .024$, $t = 3.69$, $p < .001$). An interaction for block type by session 2 was significant for response time ($\beta = -.069$, $SE = .024$, $t = 2.84$, $p = .005$), but not accuracy ($p = .75$). Other interactions for response time include block type by difficulty ($\beta = .087$, $SE = .025$, $t = 3.55$, $p < .001$) and a three-way interaction of block type, session 3, and difficulty ($\beta = -.095$, $SE = .034$, $t = 2.83$, $p < .001$).

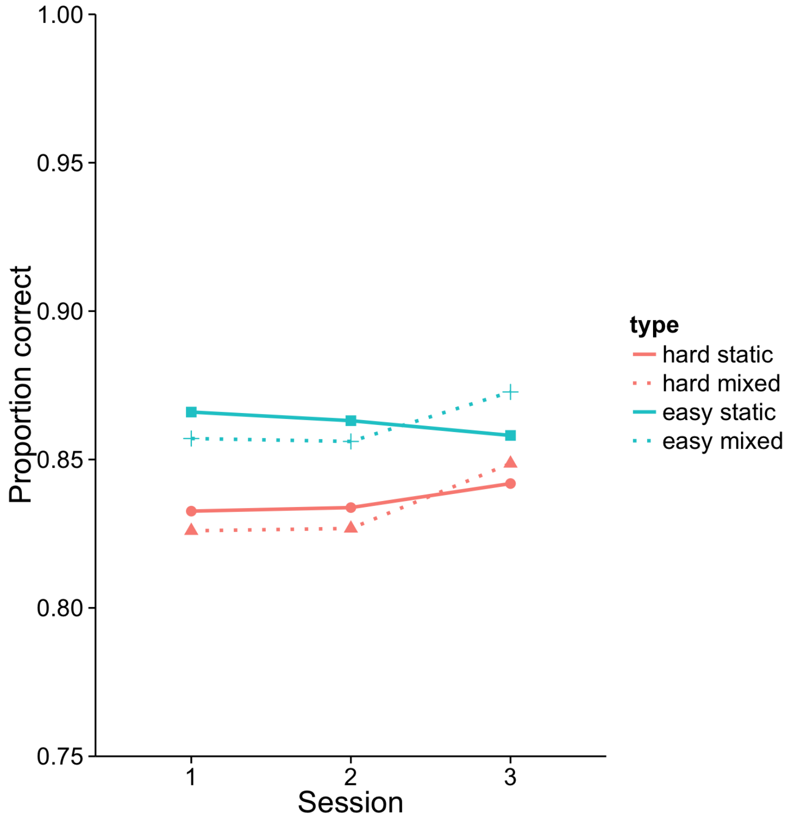

Figure 18. Accuracies initially appear higher for static compared to mixed blocks, but this pattern is not present in the third session. Note that in this graph and all subsequent accuracy graphs, the y-axis is zoomed to illustrate small effects.

Overall, there was a main effect of block type on both response time ($M_{mixed} = 0.847s$, $M_{static} = .839s$, $\beta = .069$, $SE = .022$, $t = 3.1$, $p = .002$) and accuracy ($M_{mixed} = .874$, $M_{static} = .877$, $\beta = -.02$, $SE = .008$, $t = 2.3$, $p = .02$). These are very small effects.

Despite statistical significance, all of the effects involving block type are of very small magnitude and so should be interpreted with caution. It appears that initially, accuracy is higher and response times are lower for static compared to mixed blocks. However both of these effects are absent or reversed by the third session.

Pre-cue. Pre-cue had a significant effect on accuracy ($M_{precue} = .874$, $M_{noprecue} = .876$, $\beta = -.02$, $SE = .009$, $t = 2.28$, $p = .023$), but had no effect on response

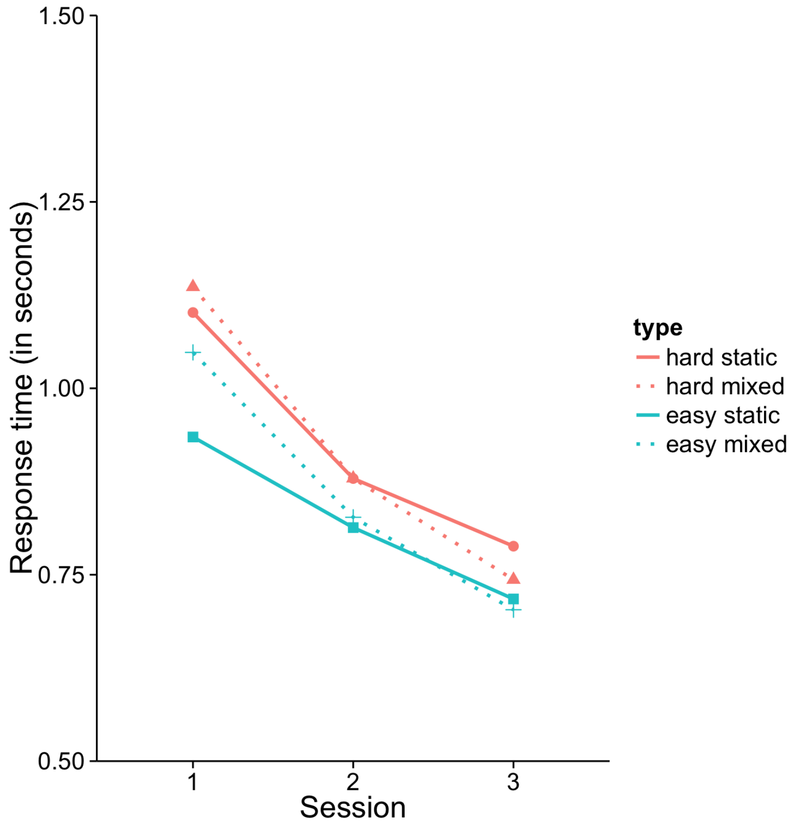

Figure 19. Response times initially appear lower for static compared to mixed blocks, but this pattern is attenuated or reversed by the third session

time ($p = .81$). This is a very small effect. Note that a main effect of pre-cue might not be expected, as half of the blocks that contained a pre-cue were of a static difficulty where this cue is not as informative. Instead, one might expect a pre-cue to have an effect on mixed blocks, but not static blocks. Indeed, there was an interaction between pre-cue and block type on accuracy ($\beta = .027$, $SE = .012$, $t = 2.32$, $p = .02$) and response time ($\beta = -.07$, $SE = .025$, $t = 2.81$, $p = .005$). See Figure 20 and Figure 21.

Additionally, there are numerous interactions involving pre-cue. A pre-cue by difficulty level interaction was significant for both accuracy ($\beta = .023$, $SE = .011$, $t = 2.04$, $p = .04$) and response time ($\beta = .049$, $SE = .024$, $t = 2.01$, $p = .045$). A pre-cue by session 3 interaction was also significant for accuracy ($\beta = .022$, $SE = .011$, $t = 1.98$, $p = .047$)

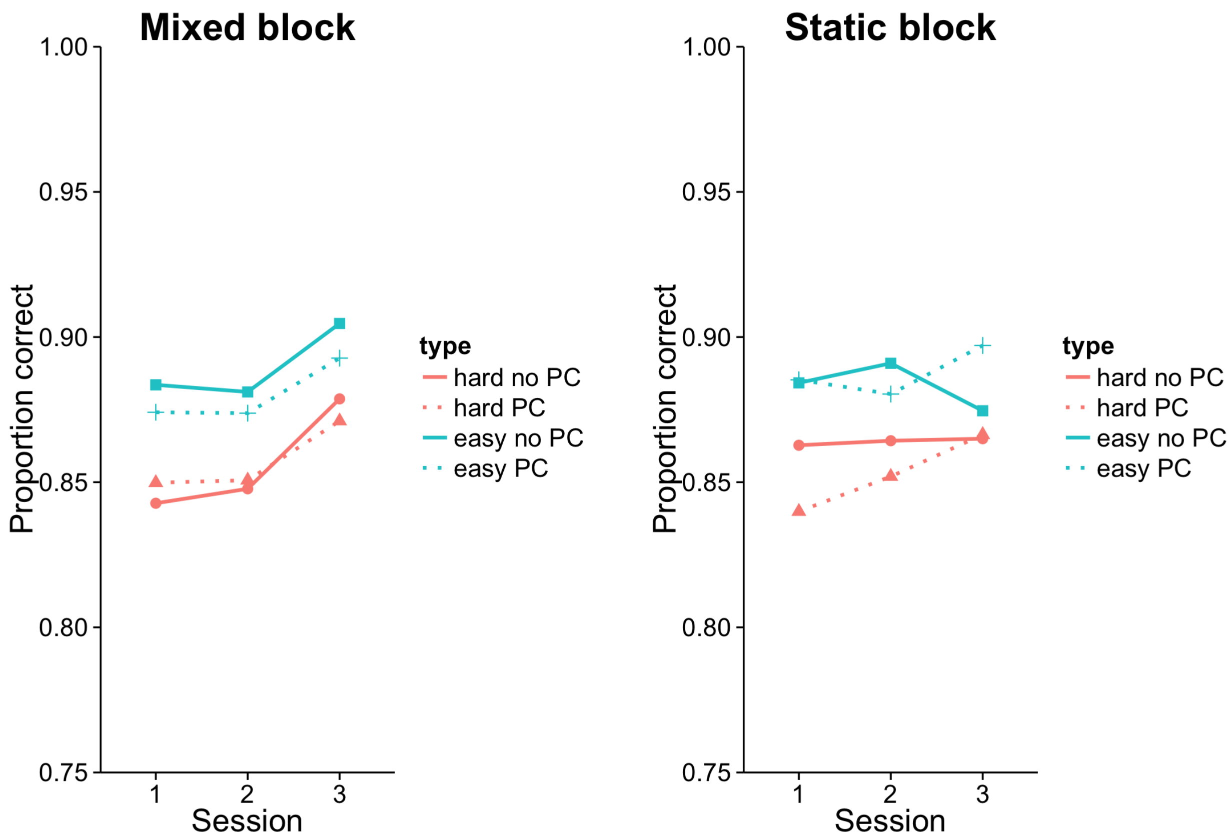

Figure 20.]

Accuracy was affected by the presence of a pre-cue, but numerous interactions made this relationship complex.

and response time ($\beta = .049$, $SE = .024$, $t = 2.05$, $p = .04$)[4].

Finally, numerous three- and four-way interactions involving pre-cue were significant. For clarity, only p-values are listed, though full statistics are available in the Appendix B. The list of interactions includes: pre-cue by block type by difficulty on accuracy ($p = .021$), pre-cue by block type by session 3 on accuracy ($p = .025$), pre-cue by difficulty by session 3 on response time ($p = .007$)[5], pre-cue by session 2 by block type on response time ($p = .027$), and pre-cue by difficulty by session 3 by block type on response time ($p = .03$)[6].

Due to the number of interactions, these effects are difficult to interpret. Generally,

[4]Not-significant with log transformed data, $p = .155$

[5]Nearly significant with log transformed data, $p = .068$

[6]Not-significant with log transformed data, $p = .236$

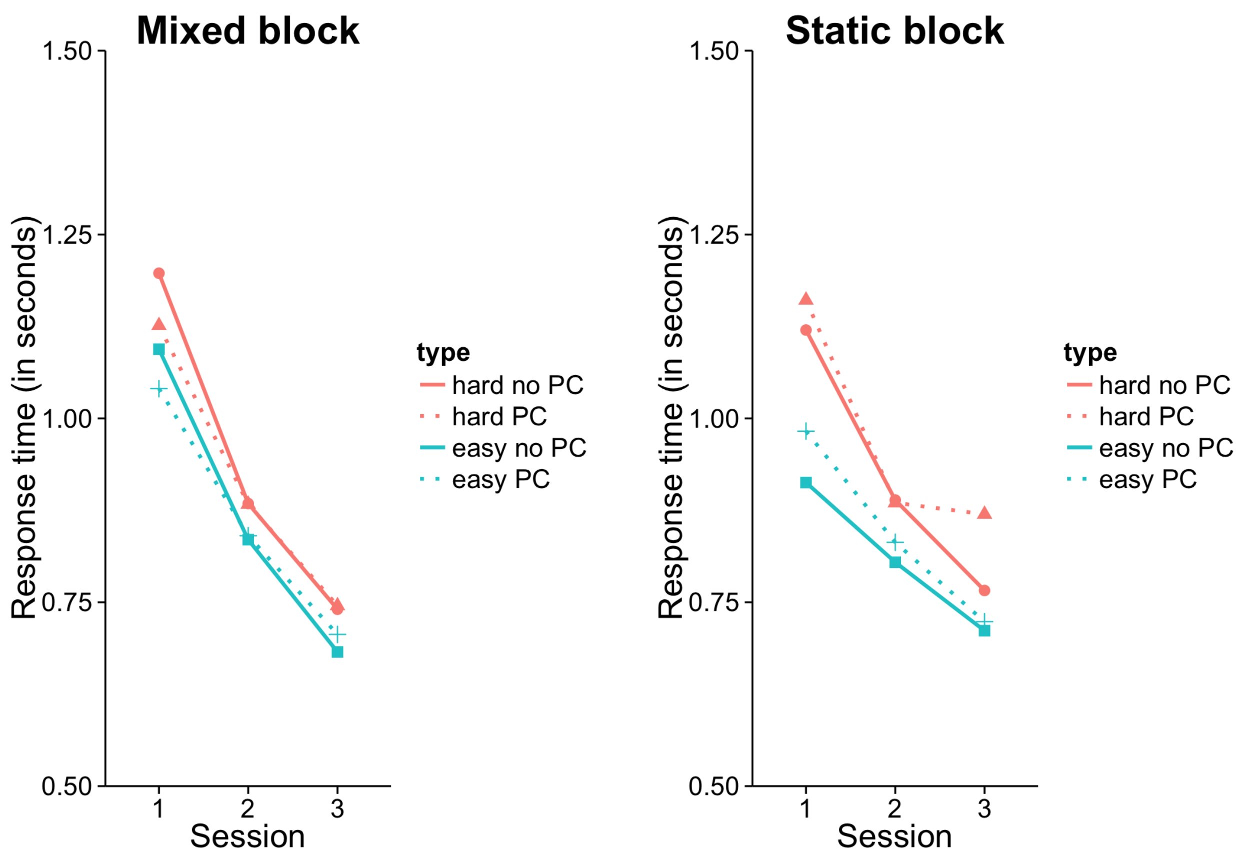

Figure 21. Response time was affected by the presence of a pre-cue, but numerous interactions made this relationship complex.

interactions of pre-cue and session could illustrate an attenuation or reversal of direction over time. In session 1, pre-cue resulted in lower response times ($M_{diff} = -0.018$) and lower accuracies ($M_{diff} = -0.004$), but in session 3 pre-cue resulted in higher response times ($M_{diff} = 0.018$) and higher accuracies ($M_{diff} = 0.003$).

Modeling Results

Diffusion model parameters (non-decision time, drift rate, and decision threshold) were derived for each subject using same methodology as Experiment 1. Separate parameters were calculated for each of cells in the design (session by block type by pre-cue by difficulty). Model parameters were then compared between conditions using a repeated-measures ANOVA.

Non-decision time. Contrary to expectations, a significant interaction of session by block type affected non-decision time, $F(2, 50) = 3.63$, $MSE = .028$, $p = .034$. Because this pattern is inconsistent across sessions (Figure 22), it is not clear what this effect means or if it is a spurious effect.

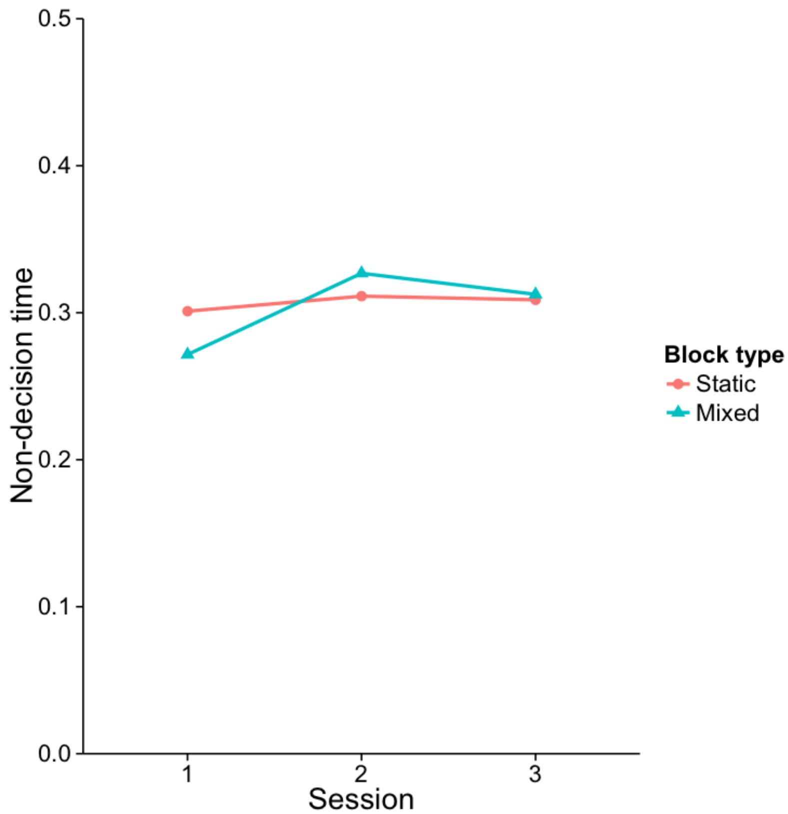

Figure 22. Static blocks appeared to have a consistent non-decision time over all sessions, whereas non-decision time in mixed blocks appeared to vary by session.

Drift rate. As expected, drift rate was significantly higher for easy ($M = .171$) compared to hard ($M = .128$) trials, $F(1, 25) = 97.69$, $MSE = .285$, $p < .001$. In addition, drift rates appeared to increase steadily across sessions, $M_{session1} = .123$, $M_{session2} = .141$, $M_{session3} = .184$, $F(2, 50) = 6.83$, $MSE = .201$, $p = .002$.

A significant session by block type interaction was found as well, $F(2, 50) = 4.63$, $MSE = .016$, $p = .014$. It appears drift rates are initially lower in mixed compared to static blocks, but this pattern is attenuated or reversed by the third session (Figure 23).

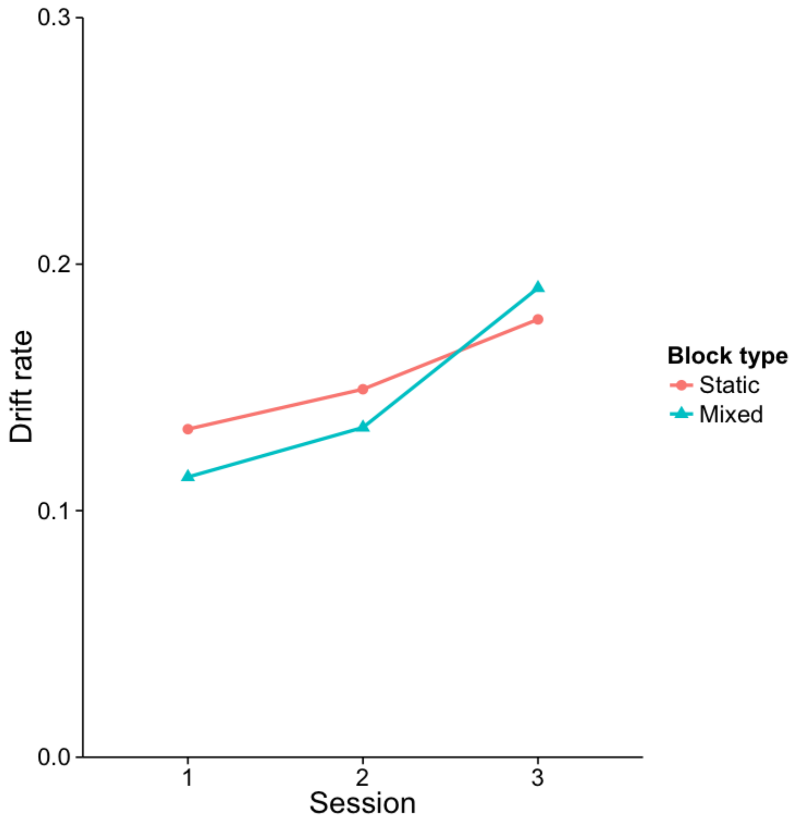

Figure 23. Drift rates were initially lower in mixed compared to static blocks, but this effect is reversed by the third session.

Decision threshold. Decision thresholds appear to decrease across sessions, $M_{session1} = .202$, $M_{session2} = .164$, $M_{session3} = .145$, $F(2, 50) = 14.75$, $MSE = 0.17$, $p < .001$. Additionally, decision thresholds were slightly higher in blocks without a pre-cue ($M_{precue} = .168$, $M_{noprecue} = .173$), however this effect was not quite significant, $F(1, 25) = 3.67$, $MSE = .003$, $p = .067$. Contrary to expectations, there were no main effects or interactions involving block type that affected decision threshold.

Optimal thresholds were computed for each session by pre-cue condition, but for the mixed blocks only, using Equation 4. This equation predicts a separate threshold for hard and easy trials. As with Experiment 1, participant thresholds were consistently set above optimal threshold levels. 78% of thresholds were above optimal ($RMSE = .1$), though this declined over the course of three sessions from 87% in session 1 ($RMSE = .149$), to 77% in session 2 ($RMSE = .08$), to 71% in session 3 ($RMSE = .067$).

It was hypothesized that participants might engage in a compensatory strategy for mixed blocks. The normative strategy outlined in Equation 4 allows one to trade off between trial types, for instance responding more quickly on hard trials in order to be presented easier trials. A non-compensatory strategy entails using a single threshold for both hard and easy trials within a block, whereas a compensatory strategy entails setting a separate threshold for each trial difficulty: one higher than the threshold predicted by the non-compensatory model, and one lower. In some sense, a higher decision threshold is an indication of preference for that trial type, though this preference may vary across subjects, session, or condition as the drift rates evolve. A qualitative test of this hypothesis was performed by examining which threshold was set higher (easy or hard trials) for each condition and comparing the predictions from Equation 4.

Results were not promising. Participant preferences for hard or easy trials matched predictions in only 75 of 156 (48%) of cases, suggesting that participants do not trade off between trial types optimally. At the group level, 90 of 156 (58%) of the predicted thresholds showed a preference for easy trials, $\chi^2(1, n = 156) = 3.69, p = .055$. Collapsing across subject and condition, observed thresholds also showed a slight preference for easy trials (85 of 156), but this result was not significant, $\chi^2(1, n = 156) = 1.126, p = .26$.

Discussion

The main research hypothesis in this experiment is that participants use a compensatory strategy in which they modulate their decision threshold based on the difficulty of other trials within the block. There did not appear to be much evidence in favor of this hypothesis. Analysis of decision thresholds found no significant main effect or interactions involving block type, suggesting participants do not modulate their decision threshold depending on the variability of trial difficulty in a block. Further, using a compensatory strategy one might expect accuracy in a mixed block to go up for one trial type and down for another relative to a static block; similarly, response time should decrease for one trial type and increase for another. Thus it was unexpected that the effects of block type are in the same direction for both hard and easy trials. That is, accuracy was initially higher for both hard and easy trials, and similarly response times were initially lower for both difficulty levels in the mixed compared to static blocks.

Surprisingly, performance in mixed blocks suffered compared to static blocks. Having ruled out changes in decision thresholds, it is not immediately clear why this is the case. If mixed blocks incurred switch costs between trial types, one might expect an overall higher non-decision time for mixed compared to static blocks, but this is not the case. The behavioral results may be due to a decrease in drift for mixed blocks, as the overall trend in drift rate matches that of accuracy and response time. Recent evidence suggests that when the prior probability of motion direction is unequal and the quality of sensory evidence is variable (i.e., blocks of mixed difficulty), drift rates are biased towards the favored prior (Hanks, Mazurek, Kiani, Hopp, & Shadlen, 2011; Moran, in press). Although trials in the current experiment had equal prior probability of leftward and rightward motion, mixed blocks may have similarly caused uncertainty about the quality in sensory evidence, leading participants to weight sensory evidence less and result in lower drift rates.

One possible explanation is that subjects have difficulty engaging in a compensatory strategy because it requires additional cognitive computations that may be costly. Specifically, in order to adjust one's decision threshold between hard and easy trials, one must be able to recognize a trial as difficult or hard prior to making a response. As response times in this task are quite short, it may be more convenient to simply adopt a single threshold for the entire block. In contrast, providing a pre-cue reduces the complexity of this computation. If participants are capable of calculating and maintaining distinct tradeoff policies for trials of different difficulties, the task is now simply akin to a lookup table: set a decision threshold based on the color of the pre-cue.

Despite a main effect and numerous interactions, it did not appear that pre-cue had a positive effect on performance, and may have even impaired performance. Perhaps presenting a pre-cue made participants feel more confident in their ability, leading them to set lower decision thresholds. Alternatively, participants may have felt an experimental demand to alter their strategy dependent on trial type; an ineffective strategy may lead to performance decrements. Anecdotal evidence supports this notion of experimental demand, as one participant commented after his third session that initially he treated trial types differently, but ultimately found it easier to ignore the pre-cue.

Overall, the very small magnitude of these effects, combined with the observation that performance decrements were absent by the third session, suggest that researchers should not be particularly worried about within-block variability of trial difficulty and pre-cues affecting performance. It is unlikely that these changes in experimental design will overshadow any intended effects from other experimental manipulations.

Conclusion

Previous research has shown a failure to optimize speed-accuracy tradeoffs in a variety of laboratory tasks (Holmes & Cohen, 2014). Many explanations for these failures have been proposed, which vary between suggestions that we optimize reward rate given limitations in our cognitive ability (e.g., Zacksenhouse et al., 2010) or that we are optimizing some quantity other than reward rate (e.g., Bogacz et al., 2010).

Experiment 1 uncovered two factors, feedback and shorter block length, that helped participants set decision thresholds closer to optimal. Additionally, feedback had a lasting effect on decision thresholds; after feedback was removed, decision thresholds changed very little. Feedback also helped improve perceptual discriminability, through an increase in drift rate.

Shorter block length also enhanced perceptual discriminability. The most likely explanation for this is a fatigue or inattention effect, in which performance decreases over the length of the block. Indeed, it is often assumed that drift is influenced not only by perceptual abilities but also attentional processes (e.g., Liu, Holmes, & Cohen, 2008). Diffusion models may be a useful and novel method for studying fatigue and vigilance, perhaps by tracking drift rate throughout a block. Future studies could use a time-shifting block (à la a moving average) to calculate diffusion parameters and generate psychometric curves to track performance over time.

This explanation may seem paradoxical in that drift rate increases with practice, yet decreases over a continuous block. After all, participants were not required to take a significant break between blocks. However, vigilance research has shown that performance in a monotonous task may decrease over time, yet remain high if one includes a single attention-grabbing trial amidst thousands (Veksler & Gray, 2008). While participants often spent no more than a minute or so between blocks, this may be sufficient to revitalize participants and provide a performance boost.

Overall, the results of Experiment 1 have important methodological implications. A plethora of popular psychological experiments allow subjects to make speed-accuracy tradeoffs (e.g., reaction time tasks, go-no/go tasks, lexical decision tasks, picture word interference tasks, implicit association tasks, random dot motion tasks, stroop tasks, cueing tasks, and signal detection tasks). Researchers should be aware of how feedback and block length can influence performance in order to compare and contrast studies that allow for these tradeoffs. Additionally, analysis of behavioral data using a sequential sampling model such as the Ratcliff diffusion model is strongly encouraged as a supplement to traditional analyses on response time and accuracy.

The results also suggest ways to improve performance in real-world tasks that allow for speed-accuracy tradeoffs, and particularly in repetitive tasks that involve signal detection, such as assembly line quality control, mail sorting, baggage screening, or chicken sexing (Biederman & Shiffrar, 1987). Workers should undergo training in an environment that provides feedback, and continue to utilize feedback on the job whenever possible. Additionally, frequent breaks are highly encouraged to sustain performance, even if these breaks are very short in duration. This is consistent with the vigilance literature, which suggests that perceptual sensitivity in a signal detection task may decline over time, but brief mental breaks are enough to combat a decline in performance (Ariga & Lleras, 2011).

Experiment 2 tested the hypothesis that people use a compensatory strategy to optimize across different trial types. While there is some evidence that participants perform differently in mixed compared to static blocks, it does not appear to be the result of optimization. If anything, mixing trial types appears to have a detrimental effect on performance, though the reasons why are still unclear. This would suggest that performance is best when participants complete a repetition of identical trials.

Like Balci et al. (2011), Experiment 2 showed no evidence of adjustments in decision threshold across blocks of varying difficulty. It is worth noting that the compensatory strategy tested in Experiment 2 is borrowed from optimal foraging theory (Charnov, 1976), in which animals decide how long to forage in a patch of food—a domain in which animals are task experts. Perhaps humans may still show evidence of a compensatory strategy in other, highly practiced decision-making tasks. It is interesting that participants in a word naming task show response times and accuracies indicative of an ability to change decision thresholds across block type (Lupker et al., 1997). Arguably, word naming is a highly practiced task compared to the more artificial random dot motion task. Future studies could examine whether performance is similar across a more diverse set of tasks.

Though participants do not show evidence of shifting decision thresholds, response time and accuracy do appear to differ depending on block type. Analysis of diffusion parameters indicated that these behavioral differences are ostensibly driven by changes in drift rate and non-decision time, though a conclusive explanation for why block type influences these parameters is still elusive. One possible limitation of the current study is that a linear mixed-effect model was used to analyze behavior, whereas a repeated-measures ANOVA was used to analyze diffusion parameters. The former is a more powerful test as it can include every trial in the regression, and thus detect very small differences in accuracy and response time. In contrast, only one set of diffusion parameters can be generated per condition, thus dramatically reducing the number of data points to be used in analysis. This negates a major benefit of using linear mixed-effect models in the first place, and thus a traditional repeated-measures ANOVA was used instead. With fewer data points the test is less powerful, thus providing an incomplete picture of the cognitive mechanisms driving behavioral changes in the study.

Yet another alternative is that the set of parameters used to model performance in Experiment 2 is simply not adequate to account for subject performance. In particular, the between-trial variability in drift rate parameter (s_v) was excluded from the model in order to calculate optimal decision thresholds; however, it is plausible that variability in drift rate is another factor influencing behavior in mixed blocks. The preceding analysis identified a drift rate and decision threshold for each difficulty level; as the coherence rate for a particular difficulty level is constant, setting between-trial variability in drift rate to zero may be an adequate assumption. If, however, trials within a block are best modeled by a single decision threshold (as observed), signal quality could instead be modeled using a single mean drift rate and variability in drift rate, rather than two discrete drift rates.

Overall, the current set of experiments directly advances our knowledge of how humans make simple binary decisions. In addition to methodological and real-world implications, the results add to a growing body of literature focused on optimality in sequential sampling models. All four proposed factors—feedback, block length, variability of trial difficulty, and foreknowledge of task difficulty—had discernible effects on subject performance. These insights provide new pieces to the puzzle that will help guide future research and modeling efforts.

References

Ariga, A., & Lleras, A. (2011). Brief and rare mental “breaks” keep you focused: Deactivation and reactivation of task goals preempt vigilance decrements. *Cognition*, *118*, 439–443.

Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Attention, Perception, & Psychophysics*, *73*, 640–657.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4 [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=lme4 (R package version 1.1-6)

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 640–645.

Bogacz, R. (2007). Optimal decision-making theories: Linking neurobiology with behaviour. *Trends in Cognitive Sciences*, *11*, 118–125.

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.

Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010). Do humans produce the speed–accuracy trade-off that maximizes reward rate? *Quarterly Journal of Experimental Psychology*, *63*, 863–891.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, *12*, 4745–4765.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Busemeyer, J. R., & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology*, *32*, 91–134.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.

Cain, M. S., Vul, E., Clark, K., & Mitroff, S. R. (2012). A Bayesian optimal foraging model of human visual search. *Psychological Science*, *23*, 1047–1054.

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*, 129–136.

Dambacher, M., & Hübner, R. (2014). Time pressure affects the efficiency of perceptual processing in decisions under conflict. *Psychological research*, 1–12.

Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An introduction to behavioural ecology.* Hoboken, NJ: John Wiley & Sons.

Donkin, C., Averell, L., Brown, S., & Heathcote, A. (2009). Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. *Behavioral Research Methods*, *41*, 1095-1110.

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review*, *16*, 1026–1036.

Estes, W., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, *12*, 403–408.

Feller, W. (1968). *An introduction to probability theory and its applications.* John Wiley & Sons.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*, 381–391.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *The Journal of Neuroscience*, *31*, 6339–6352.

Heathcote, A., Brown, S., & Mewhort, D. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, *9*, 394–401.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*, 1–10.

Holmes, P., & Cohen, J. D. (2014). Optimality and some of its discontents: Successes and shortcomings of existing models for binary decisions. *Topics in Cognitive Science*, *6*, 258–278.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York, NY: Cambridge University Press.

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the implicit association test: A diffusion-model analysis. *Journal of Personality and Social Psychology*, *93*, 353–368.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2014). lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=lmerTest (R package version 2.0-6)

Lewis, R. L., Shvartsman, M., & Singh, S. (2013). The adaptive nature of eye movements in linguistic tasks: How payoff and architecture shape speed-accuracy trade-offs. *Topics in Cognitive Science*, *5*, 581–610.

Liu, Y. S., Holmes, P., & Cohen, J. D. (2008). A neural network model of the eriksen task: Reduction, analysis, and data fitting. *Neural Computation*, *20*, 345–373.

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task:

Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 570–590.

Mackworth, N. (1948). The breakdown of vigilance durning prolonged visual search. *Quarterly Journal of Experimental Psychology*, *1*, 6–21.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco, CA: W.H. Freeman.

Moran, R. (in press). Optimal decision making in heterogeneous and biased environments. *Psychonomic Bulletin & Review*, 1–16.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*, 61–64.

Mulder, M. J., Keuken, M. C., van Maanen, L., Boekel, W., Forstmann, B. U., & Wagenmakers, E.-J. (2013). The speed and accuracy of perceptual decisions in a random-tone pitch task. *Attention, Perception, & Psychophysics*, *75*, 1048–1058.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, *7*, 308–313.

Pirolli, P. L. (2007). *Information foraging theory: Adaptive interaction with information.* New York, NY: Oxford University Press.

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://cran.r-project.org/manuals.html

Rastle, K., Kinoshita, S., Lupker, S. J., & Coltheart, M. (2003). Cross-task strategic effects. *Memory & Cognition*, *31*, 867–876.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Ratcliff, R. (2008). The EZ diffusion method: Too EZ? *Psychonomic Bulletin & Review*, *15*, 1218–1228.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*, 159–182.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356.

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*, 408–424.

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.

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: Seeing and deciding. *Proceedings of the National Academy of Sciences*, *93*, 628–633.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*, 1916–1936.

Shenoy, P., & Yu, A. (2012). Strategic impatience in Go/NoGo versus forced-choice decision-making. *Advances in Neural Information Processing Systems*, *25*, 2132–2140.

Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., & Cohen, J. D. (2009). Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1865–1897.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*, 135–168.

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.

Stephens, D. W., & Krebs, J. R. (1987). *Foraging theory*. Princeton University Press.

Straw, A. D. (2008). Vision egg: An open-source library for realtime visual stimulus generation. *Frontiers in Neuroinformatics*, *2*, 1–10.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, *14*, 1011–1026.

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, *40*, 61–72.

van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin & Review*, *16*, 987–999.

van Maanen, L., van Rijn, H., & Taatgen, N. (2012). RACE/A: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive Science*, *36*, 62–101.

Veksler, B. Z., & Gray, W. D. (2008, July). Counting sheep is a good way to get to sleep, but the occasional aardvark will wake you up: How a salient event improves performance. In V. Sloutsky, B. Love, & M. K (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 637–642). Austin, TX: Cognitive Science Societa.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*, 37–58.

Voss, A., & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology*, *52*, 1–9.

Wagenmakers, E.-J., van der Maas, H. L., Dolan, C. V., & Grasman, R. P. (2008). EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review*, *15*, 1229–1235.

Wagenmakers, E.-J., Van Der Maas, H. L., & Grasman, R. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*, 3–22.

Zacksenhouse, M., Bogacz, R., & Holmes, P. (2010). Robust versus optimal strategies for two-alternative forced choice tasks. *Journal of Mathematical Psychology*, *54*, 230–246.

Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology*, *236*, 549–573.

Appendix A

R code

All R code for linear mixed-effect models is presented in this Appendix.

Experiment 1:

```
lmer(rt ~ fb*blocklength*fborder + (1|subj) + (1|fb:subj) +
(1|blocklength:subj),data=data)

lmer(correct ~ fb*blocklength*fborder + (1|subj) + (1|fb:subj) +
(1|blocklength:subj),data=data)
```

Experiment 2:

```
lmer(rt ~ precue*coherence*session*blocktype + (1|subj) +
(1|precue:subj) + (1|coherence:subj) + (1|session:subj) +
(1|blocktype:subj),data=data)

lmer(correct ~ precue*coherence*session*blocktype + (1|subj) +
(1| precue:subj) + (1|coherence:subj) + (1|session:subj) +
(1|blocktype:subj),data=data)
```

Appendix B

Table of results

Table B1

Experiment 1: Effects on Response Time

Factor	β	Std. Error	df	t	p
(Intercept)	1.324e+00	1.636e-01	4.300e+01	8.091	3.48e-10 ***
fb	-3.418e-01	1.175e-01	3.200e+01	-2.908	0.00651 **
blocklength	2.768e-01	4.759e-02	3.600e+01	5.817	1.24e-06 ***
fborder	-1.993e-01	2.313e-01	4.300e+01	-0.861	0.39382
fb:blocklength	-2.160e-01	2.392e-02	3.758e+04	-9.032	< 2e-16 ***
fb:fborder	3.451e-01	1.662e-01	3.200e+01	2.076	0.04590 *
blocklength:fborder	-7.247e-02	6.722e-02	3.600e+01	-1.078	0.28821
fb:blocklength:fborder	2.321e-01	3.410e-02	3.758e+04	6.807	1.01e-11 ***

$.p < 07.^{*}p < .05.^{**}p < .01.^{***}p < .001.$

Table B2

Experiment 1: Effects on Accuracy

Factor	β	Std. Error	df	t	p
(Intercept)	8.105e-01	2.466e-02	4.000e+01	32.863	<2e-16 ***
fb	2.831e-02	1.471e-02	4.200e+01	1.924	0.0612 .
blocklength	-5.695e-03	8.125e-03	1.020e+02	-0.701	0.4849
fborder	1.862e-02	3.486e-02	4.000e+01	0.534	0.5963
fb:blocklength	2.315e-02	1.070e-02	3.472e+04	2.163	0.0306 *
fb:fborder	-3.897e-02	2.084e-02	4.200e+01	-1.870	0.0685 .
blocklength:fborder	-1.087e-02	1.138e-02	9.900e+01	-0.955	0.3418
fb:blocklength:fborder	-2.835e-02	1.528e-02	3.573e+04	-1.856	0.0634 .

$.p < 07.^{*}p < .05.^{**}p < .01.^{***}p < .001.$

Table B3

Experiment 1: Effects on Decision Threshold

Factor	df	SS	MSE	F	p
fb	1	0.02805	0.028048	4.696	0.0378 *
blocklength	1	0.02236	0.022357	6.994	0.0126 *
fborder	1	0.0234	0.02335	0.656	0.424
fb:blocklength	1	0.00294	0.002940	0.948	0.338
fb:fborder	1	0.02516	0.025162	4.212	0.0484 *
blocklength:fborder	1	0.00372	0.003718	1.163	0.2889
fb:blocklength:fborder	1	0.00206	0.002062	0.664	0.421

$.p < 07.^{*}p < .05.^{**}p < .01.^{***}p < .001.$

Table B4

Experiment 1: Effects on Drift Rate

Factor	df	SS	MSE	F	p
fb	1	0.02026	0.02026	12.98	0.00105 **
blocklength	1	0.002535	0.002535	3.870	0.0579 .
fborder	1	0.0088	0.008765	0.627	0.434
fb:blocklength	1	0.00216	0.0021626	2.179	0.150
fb:fborder	1	0.03172	0.03172	20.32	8.24e-05 ***
blocklength:fborder	1	0.002860	0.002860	4.367	0.0447 *
fb:blocklength:fborder	1	0.00166	0.0016566	1.669	0.206

$.p < 07.^{*}p < .05.^{**}p < .01.^{***}p < .001.$

Table B5

Experiment 1: Effects on Non-decision Time

Factor	df	SS	MSE	F	p
fb	1	0.0000	0.000006	0.000	0.987
blocklength	1	0.00257	0.002569	0.558	0.461
fborder	1	0.0235	0.02349	0.881	0.355
fb:blocklength	1	0.00133	0.001326	0.226	0.638
fb:fborder	1	0.0024	0.002449	0.107	0.745
blocklength:fborder	1	0.00414	0.004137	0.899	0.350
fb:blocklength:fborder	1	0.00292	0.002918	0.498	0.486

$.p < 07.^{*}p < .05.^{**}p < .01.^{***}p < .001.$

Table B6

Experiment 1: Effects on Deviation from Optimal Threshold

Factor	df	SS	MSE	F	p
fb	1	0.04521	0.04521	5.612	0.0240 *
blocklength	1	0.02890	0.028902	7.895	0.00839 **
fborder	1	0.0177	0.01772	0.367	0.549
fb:blocklength	1	0.00350	0.003503	0.922	0.344
fb:fborder	1	0.04511	0.04511	5.598	0.0242 *
blocklength:fborder	1	0.00221	0.002213	0.604	0.44260
fb:blocklength:fborder	1	0.00311	0.003109	0.818	0.372

$.p < 07.^{*}p < .05.^{**}p < .01.^{***}p < .001.$

Table B7

Experiment 2: Effects on Response Time

Factor	β	Std. Error	df	t	p
(Intercept)	1.108e+00	5.929e-02	6.800e+01	18.694	< 2e-16 ***
precue	4.837e-03	2.030e-02	2.460e+02	0.238	0.811836
difficulty	-1.927e-01	2.377e-02	9.200e+01	-8.109	2.18e-12 ***
session2	-2.271e-01	6.544e-02	5.600e+01	-3.471	0.001001 **
session3	-3.468e-01	6.539e-02	5.600e+01	-5.304	1.97e-06 ***
blocktype	6.873e-02	2.217e-02	1.330e+02	3.101	0.002355 **
precue:difficulty	4.883e-02	2.436e-02	7.876e+04	2.005	0.044958 *
precue:session2	-9.512e-03	2.424e-02	7.876e+04	-0.392	0.694789
precue:session3	4.937e-02	2.410e-02	7.877e+04	2.048	0.040549 *
difficulty:session2	1.097e-01	2.392e-02	7.875e+04	4.586	4.51e-06 ***
difficulty:session3	1.389e-01	2.368e-02	7.875e+04	5.866	4.48e-09 ***
precue:blocktype	-6.950e-02	2.468e-02	7.875e+04	-2.816	0.004859 **
difficulty:blocktype	8.712e-02	2.456e-02	7.876e+04	3.547	0.000389 ***
session2:blocktype	-6.856e-02	2.414e-02	7.876e+04	-2.840	0.004514 **
session3:blocktype	-8.799e-02	2.385e-02	7.876e+04	-3.690	0.000224 ***
precue:difficulty:session2	-1.463e-02	3.388e-02	7.875e+04	-0.432	0.665880
precue:difficulty:session3	-9.062e-02	3.361e-02	7.876e+04	-2.696	0.007020 **
precue:difficulty:blocktype	-3.574e-02	3.463e-02	7.876e+04	-1.032	0.30200
precue:session2:blocktype	7.544e-02	3.414e-02	7.875e+04	2.210	0.027137 *
precue:session3:blocktype	1.498e-02	3.385e-02	7.876e+04	0.443	0.658121
difficulty:session2:blocktype	-5.151e-02	3.404e-02	7.875e+04	-1.513	0.130283
difficulty:session3:blocktype	-9.532e-02	3.361e-02	7.875e+04	-2.836	0.004565 **
precue:difficulty:session2:blocktype	2.484e-03	4.807e-02	7.875e+04	0.052	0.958793
precue:difficulty:session3:blocktype	1.032e-01	4.754e-02	7.876e+04	2.171	0.029953 *

$.p < 07.^{*}p < .05.^{**}p < .01.^{***}p < .001.$

Table B8

Experiment 2: Effects on Accuracy

Factor	β	Std. Error	df	t	p
(Intercept)	8.637e-01	1.795e-02	4.800e+01	48.112	$< 2e\text{-}16^{***}$
precue	-1.955e-02	8.594e-03	1.063e+03	-2.275	0.02312^{*}
difficulty	1.999e-02	8.279e-03	1.657e+03	2.415	0.01585^{*}
session2	4.542e-04	1.527e-02	8.800e+01	0.030	0.97634
session3	3.622e-03	1.522e-02	8.700e+01	0.238	0.81248
blocktype	-2.008e-02	8.650e-03	8.310e+02	-2.322	0.02048^{*}
precue:difficulty	2.339e-02	1.146e-02	7.878e+04	2.042	0.04117^{*}
precue:session2	8.057e-03	1.140e-02	7.878e+04	0.707	0.47983
precue:session3	2.248e-02	1.133e-02	7.870e+04	1.983	0.04733^{*}
difficulty:session2	7.922e-03	1.125e-02	7.877e+04	0.704	0.48130
difficulty:session3	-1.079e-02	1.114e-02	7.875e+04	-0.969	0.33257
precue:blocktype	2.697e-02	1.161e-02	7.878e+04	2.323	0.02017^{*}
difficulty:blocktype	1.788e-02	1.155e-02	7.879e+04	1.548	0.12168
session2:blocktype	3.524e-03	1.135e-02	7.879e+04	0.310	0.75626
session3:blocktype	3.084e-02	1.121e-02	7.879e+04	2.750	0.00596^{**}
precue:difficulty:session2	-2.332e-02	1.594e-02	7.877e+04	-1.463	0.14343
precue:difficulty:session3	-5.637e-03	1.581e-02	7.878e+04	-0.357	0.72146
precue:difficulty:blocktype	-3.760e-02	1.629e-02	7.879e+04	-2.308	0.02098^{*}
precue:session2:blocktype	-1.233e-02	1.606e-02	7.877e+04	-0.768	0.44261
precue:session3:blocktype	-3.558e-02	1.592e-02	7.879e+04	-2.235	0.02543^{*}
difficulty:session2:blocktype	-1.165e-02	1.601e-02	7.878e+04	-0.728	0.46679
difficulty:session3:blocktype	1.455e-04	1.581e-02	7.878e+04	0.009	0.99266
precue:difficulty:session2:blocktype	2.876e-02	2.261e-02	7.878e+04	1.272	0.20345
precue:difficulty:session3:blocktype	1.344e-02	2.236e-02	7.878e+04	0.601	0.54772

$.p < 07.^{*}p < .05.^{**}p < .01.^{***}p < .001.$

Table B9

Experiment 2: Effects on Decision Threshold

Factor	df	SS	MSE	F	p
precue	1	0.003336	0.003336	3.673	0.0668 .
session	2	0.3469	0.17343	14.75	9.2e-06 ***
blocktype	1	0.00024	0.0002389	0.11	0.743
difficulty	1	0.00113	0.001132	0.738	0.398
precue:session	2	0.00421	0.002106	1.528	0.227
precue:blocktype	1	0.00036	0.000362	0.112	0.741
session:blocktype	2	0.00775	0.003875	1.369	0.264
precue:difficulty	1	0.00003	0.000026	0.012	0.915
session:difficulty	2	0.00257	0.001284	0.894	0.415
blocktype:difficulty	1	0.00390	0.003896	1.722	0.201
precue:session:blocktype	2	0.00051	0.0002553	0.162	0.851
precue:session:difficulty	2	0.00188	0.0009389	0.508	0.605
precue:blocktype:difficulty	1	0.00017	0.0001747	0.126	0.726
session:blocktype:difficulty	2	0.0053	0.002652	1.027	0.365
precue:session:blocktype:difficulty	2	0.00135	0.000677	0.611	0.547

$.p < 07. ^{*}p < .05. ^{**}p < .01. ^{***}p < .001.$

Table B10

Experiment 2: Effects on Drift Rate

Factor	df	SS	MSE	F	*p*
precue	1	0.00458	0.004581	1.868	0.184
session	2	0.4026	0.20131	6.829	0.00239 **
blocktype	1	0.00855	0.008553	1.844	0.187
difficulty	1	0.28460	0.28460	97.69	4.05e-10 ***
precue:session	2	0.00592	0.002962	0.876	0.423
precue:blocktype	1	0.00066	0.000660	0.178	0.676
session:blocktype	2	0.03204	0.016021	4.632	0.0143 *
precue:difficulty	1	0.00032	0.0003208	0.227	0.638
session:difficulty	2	0.00005	0.0000245	0.009	0.991
blocktype:difficulty	1	0.00077	0.0007746	0.485	0.493
precue:session:blocktype	2	0.00037	0.000187	0.049	0.952
precue:session:difficulty	2	0.00132	0.0006622	0.386	0.681
precue:blocktype:difficulty	1	0.00222	0.002223	0.989	0.33
session:blocktype:difficulty	2	0.00578	0.002892	1.182	0.315
precue:session:blocktype:difficulty	2	0.00261	0.001305	0.616	0.544

$.p < 07.^{*}p < .05.^{**}p < .01.^{***}p < .001.$

Table B11

Experiment 2: Effects on Non-decision Time

Factor	df	SS	MSE	F	*p*
precue	1	0.00077	0.000766	0.101	0.754
session	2	0.1204	0.06022	1.954	0.152
blocktype	1	0.0018	0.001802	0.257	0.617
difficulty	1	0.00146	0.001464	0.371	0.548
precue:session	2	0.00395	0.001977	0.333	0.718
precue:blocktype	1	0.00379	0.003789	0.513	0.481
session:blocktype	2	0.0568	0.028402	3.625	0.0339 *
precue:difficulty	1	0.00032	0.000317	0.058	0.811
session:difficulty	2	0.01187	0.005936	1.285	0.286
blocktype:difficulty	1	0.00131	0.001308	0.265	0.611
precue:session:blocktype	2	0.0039	0.001972	0.284	0.754
precue:session:difficulty	2	0.01249	0.006247	1.533	0.226
precue:blocktype:difficulty	1	0.00211	0.002109	0.625	0.437
session:blocktype:difficulty	2	0.0127	0.006374	0.98	0.382
precue:session:blocktype:difficulty	2	0.00344	0.001720	0.399	0.673

$.p < 07.^{*}p < .05.^{**}p < .01.^{***}p < .001.$