Essays on Timing and Economic Behavior

Thesis by
Rahul B. Bhui

In Partial Fulfillment of the Requirements for the
degree of
Computation and Neural Systems

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
Defended July 12, 2016

Rahul B. Bhui
ORCID: 0000-0002-6303-8837

In memory of Dr. Mason Bond.

ACKNOWLEDGEMENTS

The opening lines from *A Tale of Two Cities* perfectly describe graduate school:

> "It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way. . . "

For lightening the load, I thank my classmates and labmates, Ryo Adachi, Andi Bui, Yazan Billeh, Kyle Carlson, Matt Chao, Simon Dunne, Geoff Fisher, Marcelo Fernandez, Yifei Huang, Taisuke Imai, Duk Kim, Sergio Montero, Lucas Núñez, Debajyoti Ray, Welmar Rosado, Tom Ruchti, Alec Smith, Gerelt Tserenjigmid, Jay Viloria, Daw-An Wu, and Jackie Zhang. You improved my life on a daily basis.

Thanks to my family for being steadfastly present. Thanks to Biplob Ghosh for watching my blind spots, Heather Sarsons for keeping me grounded, Steven Weiss for showing me real hustle, and Romann Weber for knowing just what to say.

Thanks to Caltech, unique in its eccentricity. I met people who used lasers for mind control and who discovered new molecules in space. I heard the tale over lunch of how a preeminent prehistorian became an honorary member of a desert nomad group. I stumbled upon tenured professors battling each other while dressed as Harry Potter characters. Scientifically, its special interdisciplinary buffet kept me well fed, even if I had to eat my vegetables at times.

Thanks to Joe Henrich for getting me into the game and gently tolerating my initial (and future) foolishness. Thanks to Peter Bossaerts, Matt Shum, Bob Sherman, Ian Krajbich, John Allman, John O'Doherty, Ben Gillen, and Mike Ewens for being on my committees or otherwise providing advice, whether or not I had the sense to take it.

Finally, thanks to my advisor Colin Camerer – patient, generous, fun, and without whom I wouldn't be here. You said this was going to be a thrill ride, though neither of us knew how true that would prove to be. Thank you for teaching me the skills and strength of will necessary to hang on and start steering.

ABSTRACT

Economic activities unfold over time. How does timing influence our choices? How do we control our timing? Economic agents are considered to satisfy their *preferences* in an *optimal* fashion subject to *constraints*. Each chapter in this thesis tackles a different one of these three elements where the timing of behavior is central.

In the first chapter, I study the impact of loss aversion on preferences for labor versus leisure. In a real-effort lab experiment, I show that a worker's willingness to persevere in a task is influenced by information about task completion time. To directly assess the location and impact of reference dependence, I structurally estimate labor-leisure preferences with a novel econometric approach drawing on computational neuroscience. Once participants exceed an expectations-based reference point, their subjective values of time rise sharply, and they speed up at the cost of reduced work quality and forgone earnings.

In the second chapter, I propose and implement a method to test the optimality of individual deliberative time allocation. I also conduct experiments to study perceptual decision making in both simple decisions, where the difference in values between better and worse choices is known, and complex decisions, where this value difference is uncertain. The test reveals significant departures from optimality when task difficulty and monetary incentives are varied. However, a recently developed model based on optimality provides an improvement in fit over its predecessor.

In the third chapter, I investigate the effects of memory constraints on choice over sequentially presented options. In a study that combines experimental paradigms used to analyze memory and judgment separately, I find a close link between order effects in choice and in memory. I show that cognitive load stemming from either an externally-imposed distractor or naturally-occuring fatigue substantially weakens primacy effects. Thus disrupting memory encoding and consolidation can potentially alleviate bias in judgment.

TABLE OF CONTENTS

Chapter 1

OVERVIEW

Economic activities unfold over time. Every one of us has a limited number of moments to work, play, sleep, consume, and think. Although this is a fundamental aspect of life, much remains to be scientifically explored at the interface between time and microeconomic behavior. How does timing influence our choices? How do we control our timing? Economic agents are considered to satisfy their *preferences* in an *optimal* fashion subject to *constraints*. Each chapter in this thesis tackles a different one of these three elements where the timing of behavior is central. I conduct experiments to closely measure individual behavior, and combine tools from economics, psychology, and neuroscience to deeply analyze what results.

In Chapter 2, I study the impact of loss aversion on preferences for labor versus leisure. Taking longer than expected to complete a laborious task can cause disappointment if the expectation constitutes a reference point (Kőszegi and Rabin, 2006). This has the effect of a psychological tax on work which can be substantial in magnitude. However, the value of time and therefore the impact of reference dependence are hard to measure. When, and by how much, does loss aversion affect preferences for time use?

In a real-effort lab experiment, I show that a worker's willingness to persevere in a task is influenced by information about task completion time. To directly assess the location and impact of reference dependence, I structurally estimate labor-leisure preferences with a novel econometric approach drawing on computational neuroscience. Once participants exceed an expectations-based reference point, their subjective values of time rise sharply, and they speed up at the cost of reduced work quality and forgone earnings. Those who fall behind the reference point are demoralized as measured by ratings of task satisfaction. Moreover, the value of time rises at natural benchmarks partway to the primary reference point, indicating that reference dependence may modify behavior outside of the loss regime.

In everyday life, people allocate time on the basis of their beliefs about time expenditure. These beliefs may arise endogenously as in the case of workers with flexible hours (Camerer et al., 1997; Farber, 2015), or exogenously as when customers waiting in service queues receive delay announcements (Hassin, 1986; Guo and Zipkin, 2007). In the former setting, reference dependence can explain puzzling empirical relationships between labor supply and realized income (Kőszegi and Rabin, 2006; Crawford and Meng, 2011). In the latter setting, reference dependence may provide a reason for why customer patience adjusts to expectations of waiting time (Zohar, Mandelbaum, and Shimkin, 2002; Q. Yu, Allon, and Bassamboo, forthcoming). Strikingly, loss aversion in time can sometimes have effects qualitatively opposing those of monetary loss aversion, such as occurs in congestion pricing (Yang, Guo, and Y. Wang, 2014) because temporal and financial costs are inversely related. Which force wins out depends on the relative strength of each. My estimates suggest the former can be sizable.

In Chapter 3, I propose and implement a method to test the optimality of individual deliberation. When we choose whether to swerve out of the way when spotting a possible obstruction on the road or when we choose what product to buy at a store, we also implicitly choose when to stop processing information and actually take a course of action. Such decisions involve a tradeoff between speed and accuracy. Economic theory predicts that an optimal balance will be struck between the subjective costs and benefits of time spent deliberating versus performance attained. Do models of optimal deliberation accurately predict individual behavior?

I present a way of testing whether agents' responses to changes in the costs and benefits of deliberation are consistent with expected utility maximization. I also conduct experiments to study perceptual decision making in both simple decisions, where the difference in values between better and worse choices is known, and complex decisions, where this value difference is uncertain. The test reveals significant departures from optimality when task difficulty and monetary incentives are varied. This project includes the first test of new theoretical results by Fudenberg, Strack, and Strzalecki (2015) that characterize the optimal decision rule for environments with value uncertainty. I find that this theory fits behavior more closely than a simpler version of the commonly-used drift diffusion model from cognitive science.

If expected utility maximization appropriately describes behavior, then we can confidently construct models based on optimization and expect them to generalize well across a range of circumstances. However, if this is not the case, then alternative models must be sought to explain choices. If we moreover take expected utility as a normative criterion, then interventions such as time limits have the potential to improve welfare (Krajbich, Oud, and Fehr, 2014); for example, deadlines could prevent chronic deliberation that yields little tangible benefit when options are similar in value. My results challenge optimality on absolute descriptive grounds. However, some facets of it may improve our predictive ability, as the Fudenberg, Strack, and Strzalecki (2015) model provides an improved fit over its simpler predecessor which has the same number of parameters.

In Chapter 4, I investigate the effects of memory constraints on choice over sequentially presented options. We commonly choose from items appearing in a sequence, for example when judging a competition or evaluating multiple products pitched by a salesperson. Empirical work in various settings has found that the items appearing earliest and latest in the sequence tend to be chosen disproportionately often (e.g. Mantonakis et al., 2009; L. Page and K. Page, 2010). This parallels a long-standing body of research in the psychology of memory showing that the earliest and latest items tend to be better remembered (Ebbinghaus, 1885). If these two findings can be linked, then principles of memory should help us understand when these serial position effects will occur and what kind of interventions will bias or de-bias judgment. Can knowledge of order effects in memory guide our predictions of order effects in choice?

In a study that combines experimental paradigms used to analyze memory and judgment separately, I find a close link between order effects in choice and in memory, and observe evidence suggesting that memory causally influences choice. I show that cognitive load stemming from either an externally-imposed distractor or naturally-occurring fatigue substantially weakens primacy effects. Thus, reducing the ability or willingness of decision makers to rehearse their available options can potentially alleviate bias in judgment. Moreover, cognitive load selectively handicaps options presented early in a sequence without undermining recency effects. These results imply that effective decision making interventions could be built upon the disruption of memory encoding and consolidation.

More cognitive processing does not always reduce bias, as order effects illustrate. When one's thinking is imbalanced, thinking harder may only exacerbate the imbalance. This is a useful piece of knowledge when trying to assess what kinds of factors are conducive to rationality. My findings are also helpful for the development of interventions because while perfect memory cannot be feasibly attained, they entail that impaired memory may offer a second-best solution. The integration of memory and decision making holds many exciting practical and counterintuitive possibilities.

Chapter 2

FALLING BEHIND: TIME AND EXPECTATIONS-BASED REFERENCE DEPENDENCE

A worker may take on a task if he believes the work will be swift, and a customer may patronize a business if she anticipates prompt service and quick decision making. But if more time than expected is spent, discontentment can arise even in the absence of external penalties. In this paper, I evaluate the labor supply implications of reference points based on expected time use.

Theories of reference dependence imply that outcomes below some reference level are considered disproportionately undesirable. These theories are valuable because they enrich predictions of economic behavior in an empirically plausible fashion at the cost of minimal extra parameters. Nonetheless, to constrain their free parameters we seek a disciplined way to determine the reference point. Expectations have been increasingly studied as an attractive candidate for the reference point (e.g. Bell, 1985; Loomes and Sugden, 1986; Gul, 1991; Kőszegi and Rabin, 2006) because they help us intuitively and elegantly explain a range of empirical findings which are hard to understand using traditional assumptions (e.g. Pope and Schweitzer, 2011; Eliaz and Spiegler, 2013; Pagel, 2013; Meng, 2014; Bartling, Brandes, and Schunk, 2015). However, expectations are difficult to observe in field settings, hindering our ability to investigate their explanatory power. I run a controlled real-effort experiment allowing me to exogenously influence participants' beliefs and directly test theoretical predictions about time allocation. Furthermore, I devise a method to measure participants' time-use preferences in order to clearly perceive the impact of reference dependence.

To my knowledge, mine is the first real-effort experiment to test the theory of expectations-based reference dependence with a reference point in the time domain. Compared to other experiments that focus on reference points of money or goods, my results speak more closely to field applications that rely specifically on temporal expectations. Most notably, Kőszegi and Rabin (2006) and Crawford and Meng (2011) build models to analyze the labor supply decisions of workers with flexible daily schedules. Their models assume that workers have reference points based on rational expectations of hours worked and income attained. As I discuss below, these assumptions are crucial for capturing empirical regularities found by Camerer et al. (1997) and Farber (2005) and Farber (2008) which are inconsistent with neoclassical models as well as simpler versions of reference-dependent models. Their claims thus rest on the validity of reference points based on expected time use.

To directly establish the effect of reference dependence, I introduce a new technique enabling structural identification of individual preferences for time use. The value of time is generally harder to observe and quantify than the value of money. Money is relatively fungible, liquid, and easy to save and exchange, so its worth is more well-defined. The value of time lacks stable external benchmarks, especially when rooted in non-market activities such as leisure and when influenced by such subjective forces as reference dependence. As a result, the valuation of time poses a special challenge to the econometrician. My novel approach helps tackle this challenge, allowing me to exploit more fully the richness of my data and peer closely at individual preferences. Reference points and their effects can thus be detected and quantified through the lens of this structural model without assuming their existence *a priori*.

My experiment probes the theory from an angle that is different than normal in order to complement past experiments. These past studies have tested theoretical predictions by changing the location of the reference point. Abeler et al. (2011) do so by altering a fixed payment that their participants receive with some probability instead of their earned wage, while Ericson and Fuster (2011), A. Smith (2012), and Heffetz and List (2014) do so by modifying the probability of being able to trade endowments. Gill and Prowse's (2012) participants face different probabilities of winning as second movers in a simple sequential-move tournament. I instead change the amount of information available to participants to vary the impact of reference dependence. This type of variation also helps me to structurally estimate the loss aversion parameter, which most other studies are not suited for. Moreover, in contrast to the purely externally determined variation in other experiments, I include a belief manipulation that generates changes in expectations due to direct personal experience. This reflects how beliefs are spontaneously formed in many settings of interest, and is a source of behavioral fluctuations that theories of expectations-based reference dependence were designed to capture in the field.

Section 2.1 describes my real-effort experiment. Participants engage in many trials of a perceptual decision making task in which more time spent working leads to higher-quality choices. A tradeoff between speed and accuracy comes into play. One can spend more time, gather more sensory information, and be more likely to answer correctly and get paid. Or one can spend less time on the task and have more leisure time afterwards, but face a higher chance of being wrong and forgoing payment. I manipulate the willingness to trade accuracy for speed by providing some participants with information about how long the task will take. As further variation, one set of participants are given experience in the same task beforehand.

Section 2.2 outlines my model of reference dependence and its empirical implications. Where a person sits on the time-accuracy spectrum depends on how much they value time relative to reward. The optimal choice balances the marginal benefit of spending more time working – which reflects increased chances of winning monetary payoffs – with the marginal cost – which stems from the value of forgone leisure. A person who has reference-dependent preferences will be displeased if they spend longer than expected on the task and will take steps to avoid these negative sensations. That is, if they exceed their reference point, a psychological tax applies to each additional moment of work. The theory predicts that to mitigate these loss sensations, people will reduce time expenditure on the task. This reduction comes at the expense of accuracy, decreasing one's chances of monetary reward.

Participants who are provided common information about task completion time should hold expectations that are concentrated around that signal. According to the logic above, such participants will speed up and cut their time expenditure after missing the reference point. On the aggregate level this leads to a characteristic "piling up" of completion times as compared to the group given no such information. If, on the other hand, reference dependence is not in effect, then there should be no difference between the groups. Further, experience may lead to a kind of dynamic sophistication that changes behavior before the reference point is encountered, leading people to work at a faster pace.

Section 2.3 presents evidence in line with expectations-based reference dependence. Completion times do cluster near the exogenous reference point significantly more in the group with external information than in the group without, and this is accompanied by a decrease in their monetary payoffs. These participants finish the task at discontinuously higher rates after they pass the reference point. Those who take longer than the reference point exhibit more displeasure as measured by their ratings of task satisfaction.

Section 2.4 lays out the structural estimation of the time allocation model. In order to directly assess individual preferences, I engage in structural estimation of the time allocation model. In the economic theory, I embed econometric structure derived from a mathematical model of stochastic information processing, the drift diffusion model. The resulting structural model surmounts a problem for standard empirical approaches in the present context. Because people endogenously choose how much time to spend on each trial, we lack the exogenous variation in time expenditure needed to estimate a typical model. However, the drift diffusion model makes precise predictions about the joint distribution of time expenditure, costs, and benefits in terms of deeper parameters. This additional layer of structure enables us to infer individual preferences, and see how they change over time, and under the influence of reference dependence.

One specific empirical puzzle that concerns my results involves the labor supply of workers with flexible daily hours. Several studies of such populations have documented responses to wage changes consistent with reference-dependent preferences (Camerer et al., 1997; Chou, 2002; Fehr and Goette, 2007; Doran, 2014; Leah-Martin, 2015). In particular, people seem discontinuously more likely to stop working when they have met daily earnings targets, which in extreme cases can lead them to work more hours on low wage shifts than high wage shifts. Further, there appears to be large variation in realized earnings, which fixed earnings targets cannot account for (Camerer et al., 1997; Farber, 2005; Farber, 2008; Farber, 2015). This variation can, however, be accommodated by reference points based on rational expectations of hours worked and income attained (Kőszegi and Rabin, 2006; Crawford and Meng, 2011), which naturally fluctuate with circumstances and experience. In most of this empirical literature, reference points are either estimated as latent variables or proxied by past outcomes. So that we may more clearly observe the mechanisms at work, I exogenously vary beliefs about time use in a controlled experimental setting. My study complements especially the analysis of Crawford and Meng (2011) which invokes a structural model to address endogeneity issues. I combine experimental variation with an alternative structural approach to theirs to gauge the same central loss aversion parameter. Together these studies help triangulate the impact of reference dependence in time on labor supply decisions.

These findings also apply to the literature on service queues and delay announcements, a domain in which expectations-based reference dependence is starting to be applied. Many papers study the choice of customers to abandon service queues based on their beliefs about waiting time and the choice of firms to influence these beliefs by providing delay information (e.g. Hassin, 1986; Guo and Zipkin, 2007; Q. Yu, Allon, and Bassamboo, forthcoming). Firms walk a tightrope between departures due to customers who are not patient enough to wait as long as the announcement indicates and customers who end up having to wait longer than expected and come to disbelieve the information. My study investigates the claim that the latter group will be disappointed and consequently more likely to abandon queues when the announced time is exceeded. If valid, incorporating reference dependence may improve existing theories. For instance, Yang, Guo, and Y. Wang (2014) theoretically analyze queuing with customers who are loss averse relative to their expectations of service delay and price. They find the emergence of multiple equilibria corresponding to different patterns of expectations. They also discover that in some markets loss aversion drives a wedge between profit- and welfare-maximizing prices that would not otherwise exist. Thus reference-dependent preferences have meaningful welfare implications.

2.1 Experimental Design

The experiment was designed to facilitate precise measurement in the relevant choice dimensions of time and accuracy. Participants engaged in two blocks each of 100 trials of perceptual decision making problems. The focal part of the experiment consisted of the random dot motion task. This task is common in perception research (e.g. Newsome, Britten, and Movshon, 1989; Britten et al., 1992; Gold and Shadlen, 2007). In each trial, a hundred small moving dots are displayed in random locations. A small number of these dots ("signal" dots; 12% in the experiment) move deterministically either all left or all right, while the rest move in random directions. The participant has to choose which direction, left or right, the signal dots are moving in, and can respond at any time after the stimulus is first presented. A schematic diagram of the stimulus is displayed in Figure 2.1. Humans are able to reliably detect the correct answer under these circumstances, though some time is required to increase accuracy.[1] It is a simple task with choices that are made rather naturally, but is tedious and therefore imposes subjective costs on participants. They were paid $0.05 for each correct answer and nothing for each wrong answer. Feedback was only provided as totals at the very end of the experiment to suppress learning.

The experimental setup is depicted in Figure 2.2. Participants were divided into two conditions based on the experience they would receive across the two blocks.

[1] The computerized experiment was programmed using the Psychophysics Toolbox in Matlab. Explicit instructions explaining the task including examples were provided and the aperture (i.e. dot field) was small (side length 540 pixels) and square shaped in an attempt to minimize the chance and magnitude of training effects.

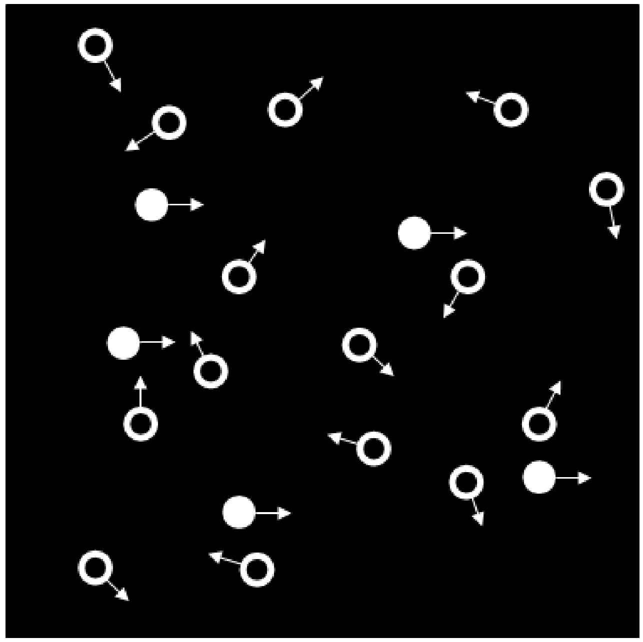

Figure 2.1: Schematic diagram of the random dot motion task stimulus

In one condition individuals completed only a single block of the focal task (random dot motion) in addition to one block of a different filler task (blurred image categorization[2]) which served to stagger their start times, avoiding confounds due to any real-time-correlated shocks. In the other condition individuals completed two blocks of the random dot motion task. In both conditions after a minute-long break between the first and second blocks were further instructions (documented in Appendix A) containing a line stating that the second block "should take about 10 minutes to complete", which constituted the experimental reference point.

Expectations-based theories rely on individual beliefs about likely outcomes. Thus, in this experiment or others, whether or not people truly believe the provided information is critical for the application of such theories. The reference point was selected because it was a natural unit of time that was feasible to attain and slightly faster than the median time expenditure in pilot tests. Although beliefs were not directly measured, there are several reasons that participants would trust the information. Given the small size of the campus, participants recruited from the Caltech population generally have firsthand or secondhand experience with the social science

[2]Participants were faced with a sequence of images of animals (raccoons and porcupines) which were obscured using standard image processing filters. For each image they chose one of the two categories. The images were from a machine learning image set collected from an online image search. They were resized to approximately 200 by 300 pixels, converted to grayscale, and obscured using a 40 pixel range Gaussian blur.

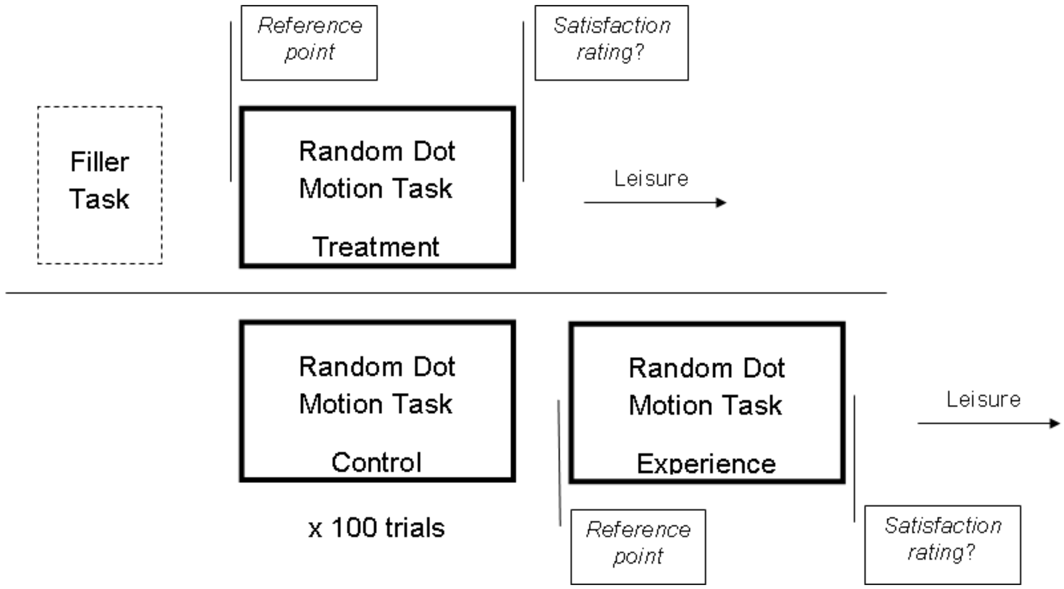

Figure 2.2: Experimental setup

laboratory. Many are used to engaging in behavioral tasks under specified financial and temporal parameters with assurances from the experimenter, the laboratory, and ultimately the institute. In line with the global prohibition of deception among experimental economists, the laboratory consent information explicitly states that "the use of deception by experimenters at SSEL is prohibited." The Caltech Honor Code also more broadly states that "no member of the Caltech community shall take unfair advantage of any other member of the Caltech community." This combination of experience and regulations should foster participant trust.

Those who faced the random dot motion task only once were given the common reference point right before that task to influence their expectations. They compose a treatment group. Those who faced the random dot motion task in the first block did so without any information and hence without a focal reference point. They compose a baseline. The second time around these same individuals thus had direct experience. They, too, were given the common informational reference point right beforehand but might respond differently than the first treatment group due to experience. Thus the presence of both external information and direct experience were varied separately, permitting us to also study the interaction between them.

During all tasks the real-time clock and trial number were displayed onscreen. At the end of the experiment participants were asked how much they liked the task on a scale of 1–10 (1 being very little and 10 being very much). They were asked to remain in their seats until at least 30 minutes had passed from the start of the experiment before being paid but were allowed to browse the internet in the meantime once they were finished. The marginal value of leisure was thus based on real leisure (Corgnet, Hernán-González, and Schniter, 2015). For laboratory timing reasons the experiment was set to end after 35 minutes and participants were informed of this.

Participants were 35 students from Caltech recruited via the online system in the Social Science Experimental Laboratory (SSEL). Seventeen people were assigned to the inexperienced condition and 18 people to the experienced condition. Two outliers from the experienced group were excluded from analysis since they took too much time and did not complete the task, and the two participants with the longest dot motion task times in the inexperienced group were also excluded to compensate. All participants received a \$5 show-up fee in addition to their earnings.

2.2 Theoretical Predictions

In this setting a person chooses the amount of time, t, to spend on each trial of the task which leads to an accuracy of $a(t)$. This accuracy function $a\colon \mathbb{R}^+ \to [0,1]$ is increasing in t, concave, and bounded, and can be interpreted as the probability of choosing the correct answer. Each correct answer yields a payoff of wage w. However, time expenditure comes at an opportunity cost of rate π. (The predictions do not qualitatively change if we instead impose the weaker assumption that the opportunity cost of time is a positive, nondecreasing, convex function $\pi(t)$.) Assuming additive separability, the expected utility function is $U(t) = wa(t) - \pi t$. The optimal choice balances the marginal benefit of increased accuracy, $wa'(t)$ (supposing differentiability), with the marginal cost of spent time, π.

People with reference-dependent preferences feel a loss when they have spent longer than the time-based reference point. When the total amount of time used has exceeded the reference point, time expenditure comes at a premium. The cost of time is then scaled up by the coefficient of loss aversion, $\lambda > 1$. Letting t_i be the time spent in each trial i, the reference-dependent utility function in trial τ is

$$U(t|r) = wa(t) - \begin{cases} \pi t & \text{if } \sum_{i=1}^{\tau} t_i < r \\ \lambda \pi t & \text{if } \sum_{i=1}^{\tau} t_i \geq r. \end{cases}$$

Figure 2.3 illustrates the effect of reference dependence assuming agents are not forward looking.[3] Before the reference point has been passed, time t_{EU} in each trial

[3]This assumption applies to agents who have limited experience. Appendix B states and explores

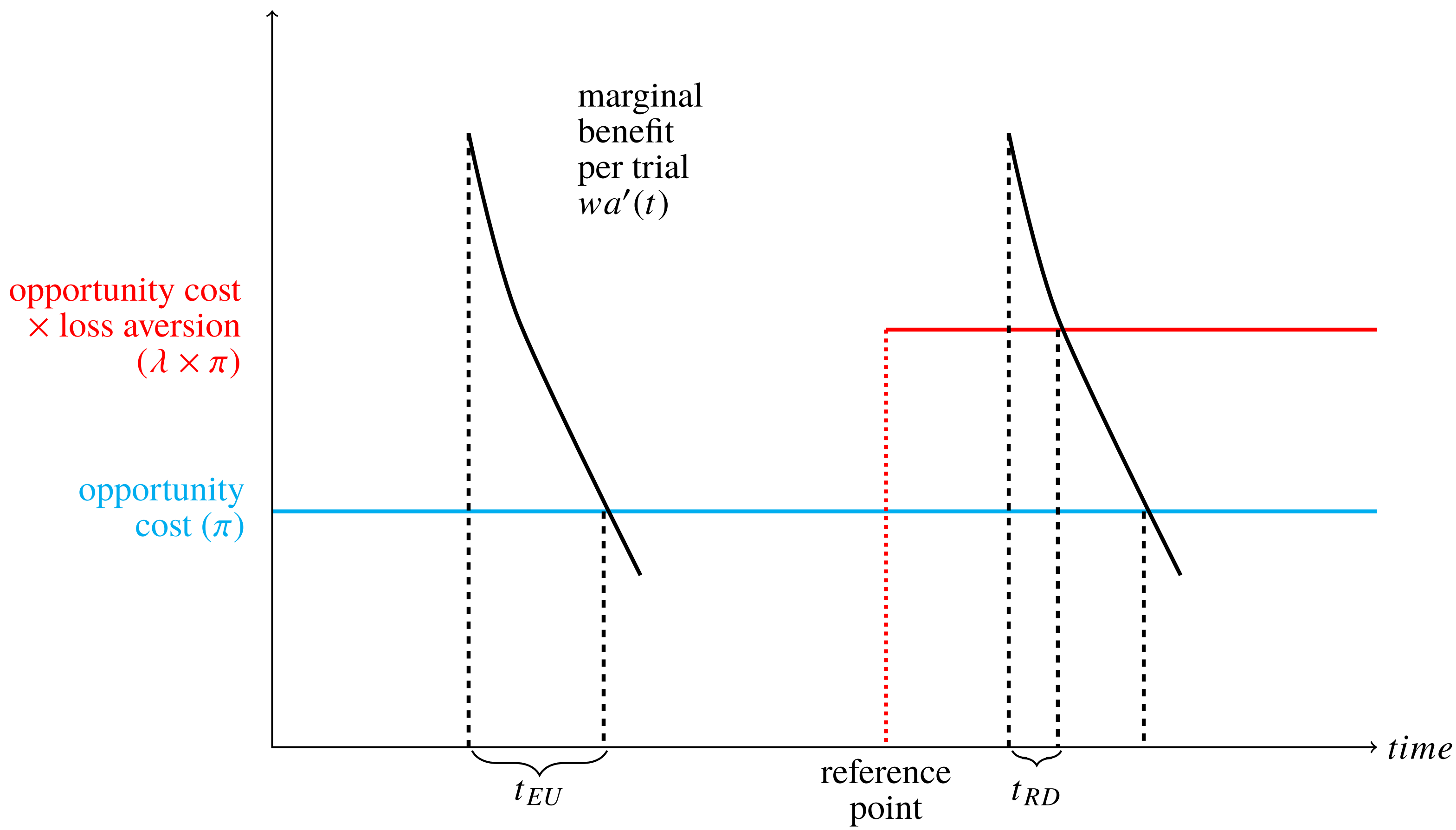

Figure 2.3: Time allocation with reference-dependent preferences

is chosen such that $wa'(t_{EU}) = \pi$, the same as a standard economic agent. Time spent in trials occurring after the reference point is penalized at a higher rate, and so time t_{RD} is spent such that $wa'(t_{RD}) = \lambda\pi$. Since $a'' < 0$ and $\lambda > 0$, this implies $t_{RD} < t_{EU}$ as is evident in Figure 2.3. It is also clear that t_{RD} is decreasing in λ, all else equal. Thus time expenditure is curtailed in trials after the reference point is hit so that people avoid incurring dramatically higher marginal costs, and the magnitude of this reduction is related to the severity of loss aversion. The drop in work time should also be accompanied by a decline in accuracy.

A signature prediction of reference dependence among a population affected by a common temporal reference point is a "piling up" of time expenditure near that point. This occurs due to the steep change in utility once the reference point is passed. For example, Allen et al. (forthcoming) report that the finishing times of marathon runners tend to bunch up at round number goals, and Markle et al. (2014) find discontinuities in satisfaction at individually-elicited marathon time goals. By a similar intuition, in my experiment, common information should lead participants' aggregate completion times to cluster near the reference point due to dissatisfaction from taking longer than expected, whereas those without such information should not fixate on this point.

2.3 Empirical Results

The data agrees with the theoretical predictions. Figure 2.4 presents the decision time distributions for the groups facing the random dot motion task for the first time. Those given the reference point appear to be clustered near it more tightly. This is confirmed by a permutation test comparing the absolute differences between time expenditure and the reference point across groups ($p = .024$), which randomly regroups the data in order to nonparametrically generate a null distribution under the hypothesis of equal mean deviations.

Result 1: *Decision times are closer to the reference point in the group given information than in the group without it, holding experience constant.*

Importantly, this difference in quitting behavior appears to be driven by the reference point. Figure 2.5 shows the empirical distribution functions depicting the cumulative probabilities of stopping across the same groups. Those given the reference point stop at a higher rate, as indicated by a Cox proportional hazards model ($HR = 2.45$,

the dynamic problem for forward-looking agents. It also explores the implications of different accuracy functions.

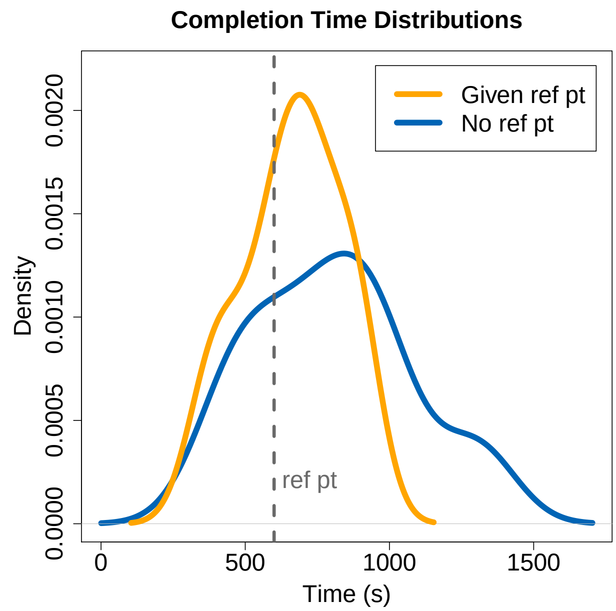

Figure 2.4: Kernel density estimate of completion time data

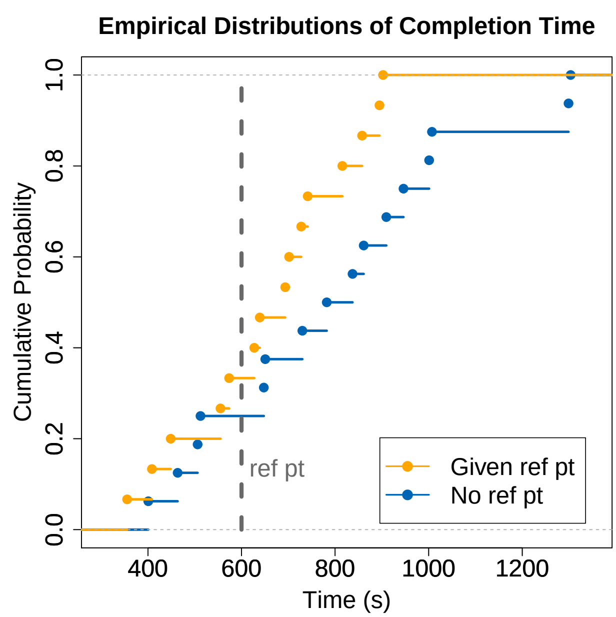

Figure 2.5: Empirical cumulative distribution function of completion time data

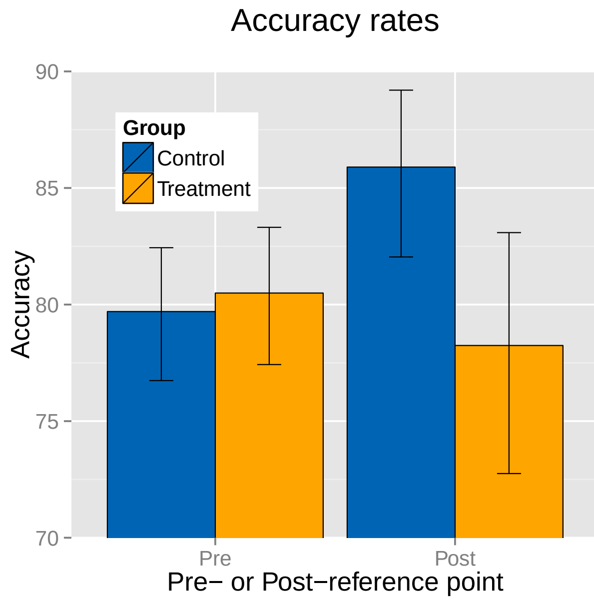

Figure 2.6: Accuracy rates across groups

p = .032). In particular, the hazard rate is significantly elevated only after the reference point is hit (HR_{after} = 3.38, p = .021; HR_{before} = 1.39, p = .620).

Result 2: *The stopping rate is higher after the reference point is hit in the group given information compared to the group without it, holding experience constant.*

Because the reference point is feasible but a little challenging to attain, participants reduce the amount of time they spend. The mean time of 663s in the group given the reference point is lower than the mean time of 804s in the group without it, which should be associated with a corresponding drop in accuracy. Figure 2.6 shows the percent of correct answers in each group for individuals active before and after the reference point, with exact 95% confidence intervals. Before the reference point, both groups responded correctly about 80% of the time. Afterwards, however, those given the reference point scored 7.7 percentage points lower than those who were not (p = .026, Z test for difference in proportions).

Result 3: *Accuracy is lower after the reference point is hit in the group given information than in the group without it, holding experience constant.*

The most direct evidence of displeasure comes from participants' post-experiment ratings of how much they liked the task on a scale from 1 to 10 with 1 being very little and 10 being very much. Figure 2.7 displays the relationship between satisfaction ratings and task completion time, with 95% nonparametric bootstrap confidence

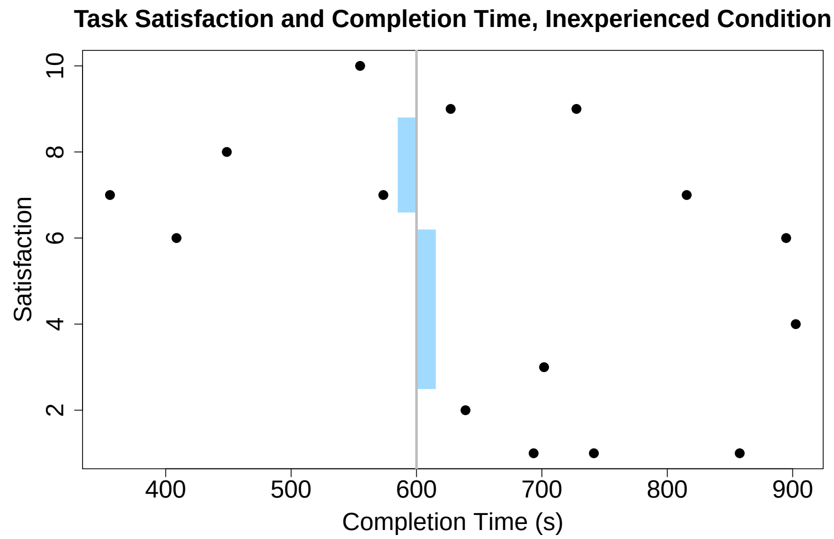

Figure 2.7: Subjective task satisfaction ratings among inexperienced participants

intervals shown. Inexperienced participants who spent longer than 10 minutes rated the task on average 3.3 points lower than those who finished quicker, a statistically significant difference according to a permutation test (p = .044).

Result 4: *Participants who are given information without experience and spend longer than the reference time on the task exhibit lower task satisfaction ratings than those who spend less time.*

With experienced participants (those who went through an extra block of the same task) the effects of training prevent the most direct comparisons of speed and accuracy from being made. Despite efforts to minimize training effects participants did seem to improve in that they completed the task more quickly but with comparable accuracy. Nevertheless, as seen in Figure 2.8 in contrast to Result 4, experienced participants who spent longer than 10 minutes rated the task a non-significant 0.5 points lower than those who finished quicker (p = .566, permutation test). Thus the experimental reference point did not appear to mark a shift in participant attitudes.

Result 5: *Participants who are given experience in addition to information and spend longer than the reference time on the task do not exhibit lower task satisfaction ratings than those who spend less time.*

Although these findings are in line with the theory, we would like to assess the impact of the treatment on preferences as directly as possible. For this, I turn to a structural approach.

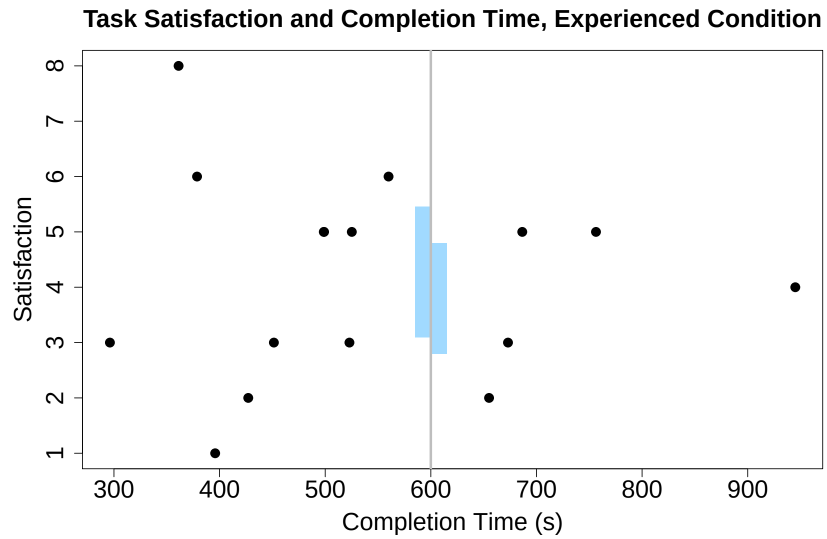

Figure 2.8: Subjective task satisfaction ratings among experienced participants

2.4 Structural Analysis

The value of time is subjective and not directly observable in this setting. This poses a challenge to our ability to quantify how loss aversion affects the value of time. We have one constraint – the first-order condition – to guide our inference. Since we observe time choices, identification of the opportunity cost of time rests on our ability to estimate the marginal benefit of working. This benefit stems from the connection between time and accuracy since the more time is spent accumulating information, the higher are one's chances of answering correctly and earning a payoff.

To estimate the marginal benefit of time, a few possibilities naturally suggest themselves. One could try to predict accuracy from decision time across an individual's trials by running a logistic, polynomial, or nonparametric regression. Some of these have obvious issues here; a polynomial regression is clearly misspecified since the fitted model will predict probabilities outside the [0, 1] bounds and will have derivatives that grow extreme or fluctuate noisily in sign, and nonparametric techniques are known to perform poorly when applied to the derivatives with which we are primarily concerned. However, all of these methods suffer from a more serious flaw in the present setting: the link between time and accuracy is hidden due to endogeneity of the decision rule. Because people choose of their own accord when to stop accumulating information, we do not have the exogenous variation that would be required to estimate how their time expenditure translates into performance. If participants were exogenously forced to stop and give their best guess at various points in time (in what is known as an "interrogation" paradigm), they would indeed be more accurate when stopped after more time. But if participants have control over when to stop (as in the "free response" paradigm I use), this relationship no longer holds. Any superficial correlation does not reflect their deeper connection, even one that is ostensibly positive.

If we were to persist with a naïvely direct approach, we would find that the amount of time people spend on each trial appears largely unrelated to their earnings. Logistic regressions attempting to predict accuracy from time for each participant in each group hold almost no predictive power. For these individual-level regressions, Figure 2.9 shows the histogram of likelihood ratio statistics that assess the goodness of fit of the model including time as a predictor as compared to the null model. Figure 2.10 shows the histogram of Z statistics that assess the statistical significance of the time coefficient in each logistic regression. Including time in the regression model yields no statistically significant benefit in 85% of cases according to these model and coefficient significance tests. In only a single case is the coefficient on time positive and statistically significant. Attempts to directly estimate the accuracy function thus seem ill-fated.

I follow an alternative route and use clues from psychology and neuroscience to capture the data generating process more fully. The random dot motion task at the center of this experiment is used often in perception research (e.g. Newsome, Britten, and Movshon, 1989; Britten et al., 1992). Patterns of choice, response times, and neural activity in this kind of setup are mathematically well-described by the drift diffusion model, in which information is accumulated with noise until confidence in one or another answer reaches a threshold, at which point the choice is made (Gold and Shadlen, 2007). This model provides a precise statistical account of how deeper cognitive parameters give rise to time spent and performance attained, and in so doing, allows for a richer interpretation of the same data.

The drift diffusion model (DDM) is a neurally plausible descriptive model of decision making in discrete choice situations. It was originally developed half a century ago (Stone, 1960; Laming, 1968; Ratcliff, 1978) and is the earliest and most well-characterized theory in a class of stochastic accumulation models that predict the joint distribution of choice probabilities and response times (Busemeyer and Townsend, 1993; LaBerge, 1962; Pike, 1966; Vickers, 1970; Usher and McClelland, 2001; Shadlen and Newsome, 1996; X.-J. Wang, 2002). Beyond behavioral

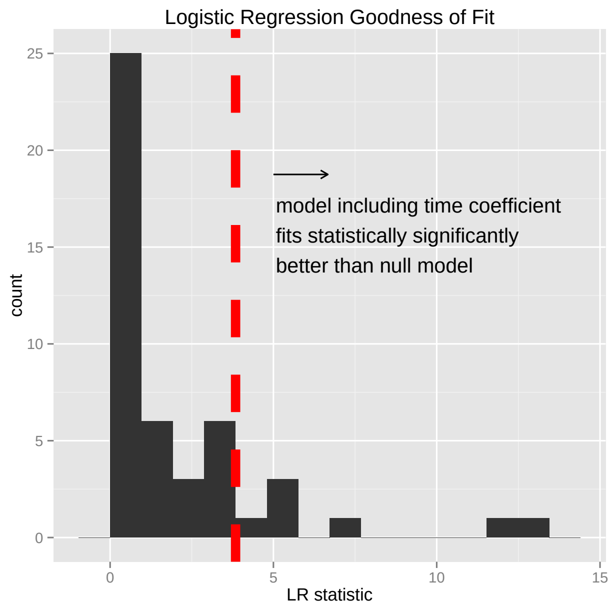

Figure 2.9: Goodness of fit of logistic regressions predicting accuracy from time expenditure

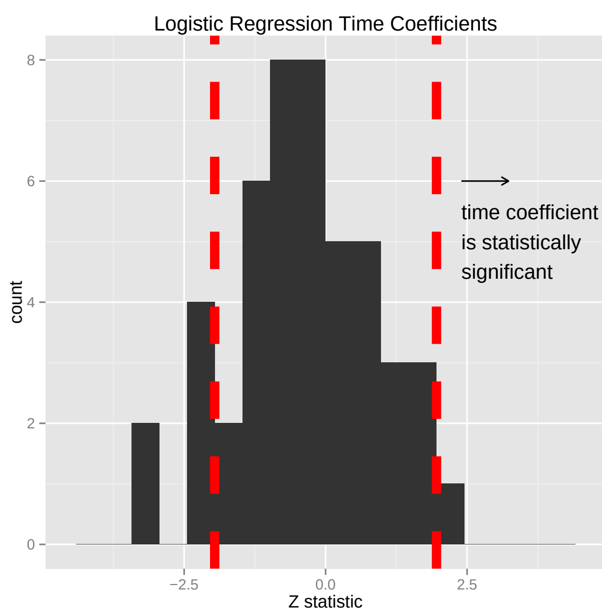

Figure 2.10: Z-scores of time coefficients in logistic regressions predicting accuracy from time

evidence showing that the DDM closely fits patterns of choice and response times in a variety of decision tasks, direct recordings of neural activity demonstrate that neurons in various brain regions implement evidence accumulation processes that match the model's structure (Gold and Shadlen, 2002; Hanes and Schall, 1996; Shadlen and Newsome, 2001; Ratcliff, Cherian, and Segraves, 2003; P. L. Smith and Ratcliff, 2004). Indeed, the basic functioning of neurons involves transmitting all-or-nothing signals that are triggered by inputs reaching a critical threshold. While the DDM is commonly used to study perceptual choice (Ratcliff and Rouder, 1998; Ratcliff, Cherian, and Segraves, 2003; Ratcliff and P. L. Smith, 2004; P. L. Smith and Ratcliff, 2004; A. Voss, Rothermund, and J. Voss, 2004; Philiastides, Ratcliff, and Sajda, 2006; Gold and Shadlen, 2007; Ratcliff, Philiastides, and Sajda, 2009), recent work extends it to value-based settings such as consumer purchasing decisions and intertemporal choice (Krajbich, Armel, and Rangel, 2010; Milosavljevic et al., 2010; Krajbich and Rangel, 2011; Krajbich, Lu, et al., 2012).

According to the DDM, the agent integrates evidence over time for one alternative or another until an evidence threshold is reached, and the corresponding decision is made. This accumulation includes inherent sensory noise and hence is modeled as a stochastic differential equation,

$$dx = Adt + cdW,$$

where $x(t)$ is the difference in evidence between the two alternatives (with $x(0) = 0$ in an unbiased decision), A is the accumulation or drift rate, and c represents the noise component.[4] The change dx over the small time interval dt is broken up into the constant drift Adt and the Gaussian white noise cdW with mean 0 and variance c^2dt. When the accumulated evidence x reaches the critical threshold $\pm z$, the corresponding choice is made.

This process generates a speed-accuracy tradeoff governed by the confidence threshold (z). A higher threshold entails a more stringent standard of evidence and reduced susceptibility to errors at the cost of greater decision time. Conversely, a lower threshold requires weaker evidence and thus less time to make a decision but increases the error rate. The drift rate (A) and noise (c) parameters describe an individual's information processing faculties, and in this economic context can be

[4]The DDM can be formulated equivalently in Bayesian terms (e.g. Arrow, Blackwell, and Girshick, 1949; Bitzer et al., 2014), and is optimal in the sense that it is the continuous sampling limit of the sequential probability ratio test, which minimizes response time for a given error rate in a Bayes optimal way.

interpreted as measures of worker ability. Higher drift rates and lower accumulation noise mean superior performance in terms of higher accuracy rates with the same threshold. These parameters can be estimated by methods known in psychometrics, one of which is given below.

We can refine our view of the endogeneity problem in light of this model. People choose when to stop accumulating evidence based partly on their belief about the state of the world (in this task, the direction of dot motion). They respond when they are sufficiently confident in the quality of their answer. However, this confidence level is also statistically related to the accuracy of their response. Without a measure of the beliefs that give rise simultaneously to time and accuracy, we are afflicted by endogeneity. In fact, the simple version of the DDM (which applies here) implies that observed speed and accuracy are superficially independent (e.g. Stone, 1960; Fudenberg, Strack, and Strzalecki, 2015), which explains the null pattern we see in the data as displayed in Figures 2.9 and 2.10. Intuitively, this happens because every decision is triggered by the same level of confidence, irrespective of how much time was taken to reach that point. As a result, conditioned on the fact that a decision was made, time expenditure does not carry any additional information with which to predict accuracy.[5] The structure of the DDM allows us to infer the confidence threshold and provides us with the measurement of beliefs we can use to address endogeneity.

I recast the decision problem in terms of the DDM parameters to exploit its underlying structure. While estimating the benefit curve as a function of time is problematic, estimating the benefit curve as a function of the decision threshold turns out to be feasible. Given our mathematical understanding of the DDM, the entire utility function can be formulated with the decision threshold substituting for time as the choice variable. This adapted utility function can be optimized as normal to obtain a first-order condition. The resulting condition is enough to infer the economic costs and benefits from the DDM parameters, which can be estimated in a first stage using methods from psychometrics.

[5] Choice patterns represented by the model can potentially exhibit a positive, neutral, or negative relationship between time and accuracy, depending on the shapes of the decision threshold and cost function (Fudenberg, Strack, and Strzalecki, 2015). In a commonly used extension (Ratcliff, 1978), response times could be *higher* on incorrect trials, inducing a *negative* relationship between time and accuracy. For instance, suppose some trials happen to be subjectively harder than others. Prolonged responses would indicate hard trials which are not worth spending too much time on and are thus less likely to be answered correctly. If anything, we see minor indications of this negative association in the logistic regressions as shown in Figure 2.10.

Key mathematical properties of individual performance conditional on the DDM parameters have been characterized, including the accuracy as a function of the decision threshold. These properties come from solutions to the first passage problem in which the stochastic accumulation process crosses the decision threshold. In the current simple setup,[6] a closed-form expression exists for the error rate, ER (e.g. Ratcliff, 1978; Rafal Bogacz, E. Brown, et al., 2006):

$$ER = \frac{1}{1 + e^{2Az/c^2}}.$$

The accuracy function we want to estimate is simply the complement of this error rate. Accuracy is thus a logistic function of the threshold, rather than a function of time directly. When multiplied by the payoff for a correct answer, this yields the benefit curve.

To fully rework the utility function, we also need to revise its cost segment, requiring a definition of time expenditure in terms of the DDM parameters. In addition to the error rate, we make use of the closed-form expression for the mean decision time as well, DT (e.g. Ratcliff, 1978; Rafal Bogacz, E. Brown, et al., 2006):

$$DT = \frac{z}{A} \tanh\left(\frac{Az}{c^2}\right).$$

We can combine the DDM with the economic model presented earlier by using the latter as a shell and appealing to the DDM to detail the functional forms generating accuracy and time. The decision threshold becomes the basic dimension of choice and time becomes implicit.

$$\begin{aligned} U(z|r) &= wa(z) - \begin{cases} \pi t(z) & \text{before reference point} \\ \lambda \pi t(z) & \text{after reference point} \end{cases} \\ &= w(1 - ER) - \pi DT \times \begin{cases} 1 & \text{before reference point} \\ \lambda & \text{after reference point} \end{cases} \end{aligned}$$

$$U(z|r) = w\left(\frac{e^{2Az/c^2}}{1 + e^{2Az/c^2}}\right) - \pi\left(\frac{z}{A}\tanh\left(\frac{Az}{c^2}\right)\right) \times \begin{cases} 1 & \text{before reference point} \\ \lambda & \text{after reference point.} \end{cases}$$

A subtle assumption is being made here. The DDM supposes a constant decision threshold within trials. While there is some contention surrounding this property,

[6] Additional parameters dealing with, for example, variation in drift rate across trials are sometimes incorporated in the extended DDM.

it is justifiable on both theoretical and empirical grounds. Theoretically, a constant threshold is indeed optimal in the present task for a Bayesian decision maker (e.g. Shiryaev, 1969; Fudenberg, Strack, and Strzalecki, 2015). Empirically, the cross-paradigm reanalysis of Hawkins et al. (2015) find evidence primarily in favor of a fixed threshold.

The standard optimization criterion is equivalent to the Bayes Risk criterion developed by Abraham Wald and Wolfowitz (1948) and Edwards (1965) which assumes decision makers minimize the cost function $BR = c_1 DT + c_2 ER$ and is known to have a unique solution. The first-order condition with respect to z yields

$$\frac{\partial U}{\partial z} = 0 = w\underbrace{\left(\frac{2Ae^{2Az^*/c^2}}{c^2(1+e^{2Az^*/c^2})}\right)}_{\phi(z^*,A,c)} - \underbrace{\left[\frac{z^*}{c^2}\operatorname{sech}\left(\frac{Az^*}{c^2}\right) + \frac{1}{A}\tanh\left(\frac{Az^*}{c^2}\right)\right]}_{\psi(z^*,A,c)} \pi \times \begin{cases} 1 & \text{before r.p.} \\ \lambda & \text{after r.p.} \end{cases}$$

$$\hat{\lambda\pi} = \frac{w\phi(z^*, A, c)}{\psi(z^*, A, c)} = \frac{wA^2}{2Az^* + c^2 \sinh(2Az^*/c^2)}. \tag{$*$}$$

In this way the opportunity cost of time is identified for each person from the decision threshold (z), drift rate (A), and accumulation noise (c) parameters which can be estimated from individual accuracy and response time data.

This indicates a two-step procedure for estimating opportunity costs on an individual level, which can be subsequently compared across treatments to estimate the loss aversion parameter. First, the DDM parameters for each person are estimated with any of several techniques regularly used in the psychometrics literature, and second, the resulting parameter values are plugged into the rearranged first-order condition ($*$) to back out each person's opportunity cost.

In the first step, for simplicity I use the EZ-diffusion model approach to estimate the DDM parameters (Wagenmakers, Van Der Maas, and Grasman, 2007). This entails closed-form solutions for the parameters based only on the proportion of correct decisions (P) and the variance in response times for correct decisions (VRT). We can see indications here of the DDM's greater interpretive power in that the procedure makes use of the variance in decision times rather than simply the mean, and does so in an intricate nonlinear fashion. The drift rate A and decision threshold z are given by

$$A = \operatorname{sign}\left(P - \frac{1}{2}\right) c \left\{\frac{\operatorname{logit}(P)\left[P^2\operatorname{logit}(P) - P\operatorname{logit}(P) + P - \frac{1}{2}\right]}{VRT}\right\}^{\frac{1}{4}}$$

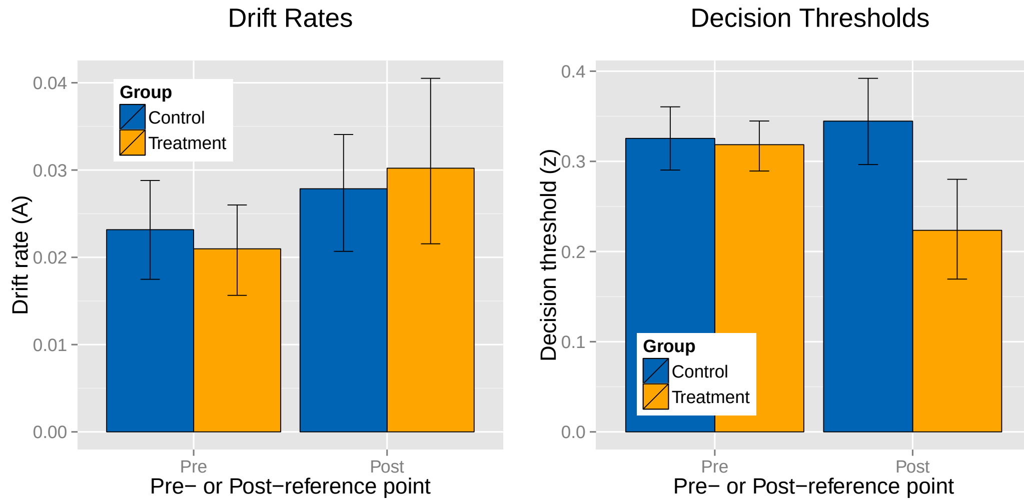

Figure 2.11: Mean estimated DDM parameters

$$z = \frac{2c^2 \text{logit}(P)}{A}.$$

where $\text{logit}(P) = \log(P/1 - P)$. The properties of the DDM depend only on the ratios z/c and A/c rather than their absolute values so $c = .1$ is assumed in estimation as is standard practice.

The resulting estimates from this step are shown in Figure 2.11 with 95% non-parametric bootstrap confidence intervals. The ability parameter is matched across groups, though it exhibits a mild increase likely due to improvement from experience. However, the decision threshold drops after the reference point is hit for the group provided with the information ($p = .001$, across-group permutation test). This structurally entails a reduction in time spent at the expense of accuracy. Since the decision threshold is considered the choice variable, this gives us some confidence that the components of the modeled mechanism are moving as they should.

In the second step, as we have seen, the utility function includes accuracy as a function of the DDM parameters. The first-order condition gives us the optimal decision threshold conditional on the other parameters. It provides an estimating equation that links all of the parameters together. We can then plug in our estimates of the parameters obtained in the first stage to obtain the remaining unknown value: the value of time.

I use this procedure to estimate the value of time for participants in both groups based on their trials before and after the reference point was passed. Thus the people represented in each group are the same across periods in this analysis. To

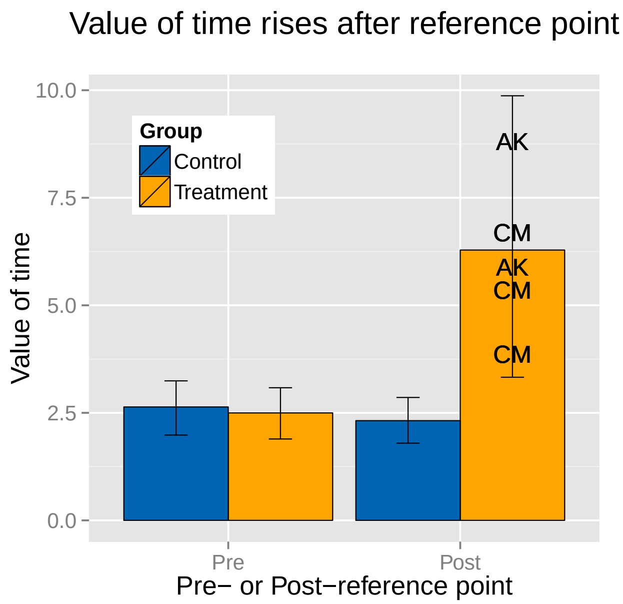

Figure 2.12: Mean estimated values of time

keep noise low in these estimates, I include participants who spent enough time to face at least 20 trials after the reference point came into effect. The values resulting from the procedure on a $\$/hr$ scale are shown in Figure 2.12 with 95% nonparametric bootstrap confidence intervals. The numbers may be somewhat lower than anticipated, which happens because only the most direct financial motivation, the piece rate payment, is accounted for. Any additional motivation for work, such as an intrinsic desire for success, would imply higher values.

Table 2.1 contains the results of regressions predicting the values based on period (pre- vs post-reference point, coded as 0 and 1 respectively) and group (no-information control vs reference point information treatment, coded as 0 and 1 respectively). As is apparent from the figure as well as the significant positive interaction between the two variables, the value of time rose dramatically only among those who were provided information and only after they exceeded the reference point. Before the reference point was passed, the value of time was the same regardless of whether groups were provided information. After the reference point was passed, the control group remained the same, passing a placebo test.

We can compare opportunity costs both between groups and within individuals to estimate the strength of loss aversion. Under the model as specified in Section 2.2, the loss aversion parameter λ is given as the ratio between opportunity costs

Table 2.1: Effect of reference dependence on value of time

	Dependent variable:	
	Value of Time	
	(1)	(2)
Constant	2.636***	3.187
	(0.782)	(1.892)
Treatment	−0.138	−2.149
	(1.219)	(2.702)
Period	−0.365	−0.365
	(1.106)	(1.141)
Treatment × Period	4.388**	4.388**
	(1.723)	(1.778)
Individual Fixed Effects	No	Yes
Observations	34	34
R^2	0.334	0.646
Note:	*p<0.1; **p<0.05; ***p<0.01	

when loss aversion is and is not in effect. Note that because the value of a correct response factors in multiplicatively as seen in ($*$), this is robust to assumptions about the utility of winning, including heterogeneous risk attitudes or psychological success bonuses. The within-person estimate is based on the ratio of post-reference point to pre-reference point opportunity costs for each individual in the treatment group. The between-group estimate is based on the ratio of treatment to control group opportunity costs after the reference point. Although we cannot observe the post-reference point behavior of individuals who finish the task too quickly, if the loss aversion parameter is independent of the baseline opportunity cost, these assessments will provide unbiased estimates of loss aversion in the population. They agree with each other reasonably well. The mean within-person estimate is $\hat{\lambda}_W = 3.330$ with a 95% nonparametric bootstrap confidence interval of $[1.337, 6.410]$, and the mean between-group estimate is $\hat{\lambda}_B = 2.871$ with a confidence interval of $[1.450, 4.939]$.

These quantities summarize how intensely shortfalls of time are felt compared to surpluses. The figures here appear similar in magnitude to the most comparable figures from the few other time-based studies that exist. The best fitting specifications of Crawford and Meng's (2011) structural analysis yield values of 1.671, 2.309, and

2.886, and Abdellaoui and Kemel's (2014) find mean values of 2.54 and 3.80.[7] To visually compare these figures, I calculate the mean post-refererence-point values of time that would result from each estimate of loss aversion, and plot the values in Figure 2.12, denoted as CM and AK. Crawford and Meng's numbers may be smaller because they examine loss aversion in time and money simultaneously. Notably, my estimates are at least as strong as commonly cited numbers reflecting loss aversion in the monetary domain. Thus we see quantitative evidence beginning to converge on the strength of loss aversion in the time dimension.

The above analysis was conducted with a particular reference point in mind. However, the technique discussed can be used to not only measure the strength of loss aversion, but also to detect the location of reference dependence. Rather than dividing the data into pre- and post-reference point, I estimate time-use preferences in a sliding window. Displayed in Figure 2.13 is the 10% trimmed mean of the value of time in each group based on a 3-minute window for the period of time in which at least 3 individuals were active.[8] The inferred preferences are relatively stable and comparable in both groups until the reference point is hit. At that point, the mean value of time among the remaining participants given the reference point sharply rises. The bar above the data denotes a statistically significant difference in group means at the 10% level according to a permutation test. Thus reference points can be identified from the data itself. Intriguingly, if I shrink the data window to 2 minutes, which increases the signal at the expense of noise, additional points of interest are revealed, shown in Figure 2.14. In particular, jumps are observed approximately halfway to and one minute before the reference point. Though not proposed beforehand, the location of these jumps suggests that they may be benchmark effects. These timings are natural markers that could be used to set one's pace, triggering changes in speed for those who feel they are behind.

An agent who takes advantage of benchmarks is behaving in a sophisticated fashion. Although there are hints of this occurring among inexperienced participants, individuals who have experience in the time allocation task may be better at pacing themselves. Figure 2.15 shows the distribution of completion times among

[7] Note that Crawford and Meng's (2011) parameter includes a coefficient that reflects the strength of reference dependence, and Abdellaoui and Kemel's (2014) study involves framed gambles over amounts of time that participants were made to spend in a room later without entertainment, and are therefore not perfectly comparable to my estimates. De Borger and Fosgerau (2008) quantitatively assess loss aversion from hypothetical travel time choices but do not estimate a comparable parameter.

[8] I exclude each individual's 100th trial and maximum estimated value of time, due to a sizable drop in performance specific to the last trial, and to ensure the results are not driven by other outliers.

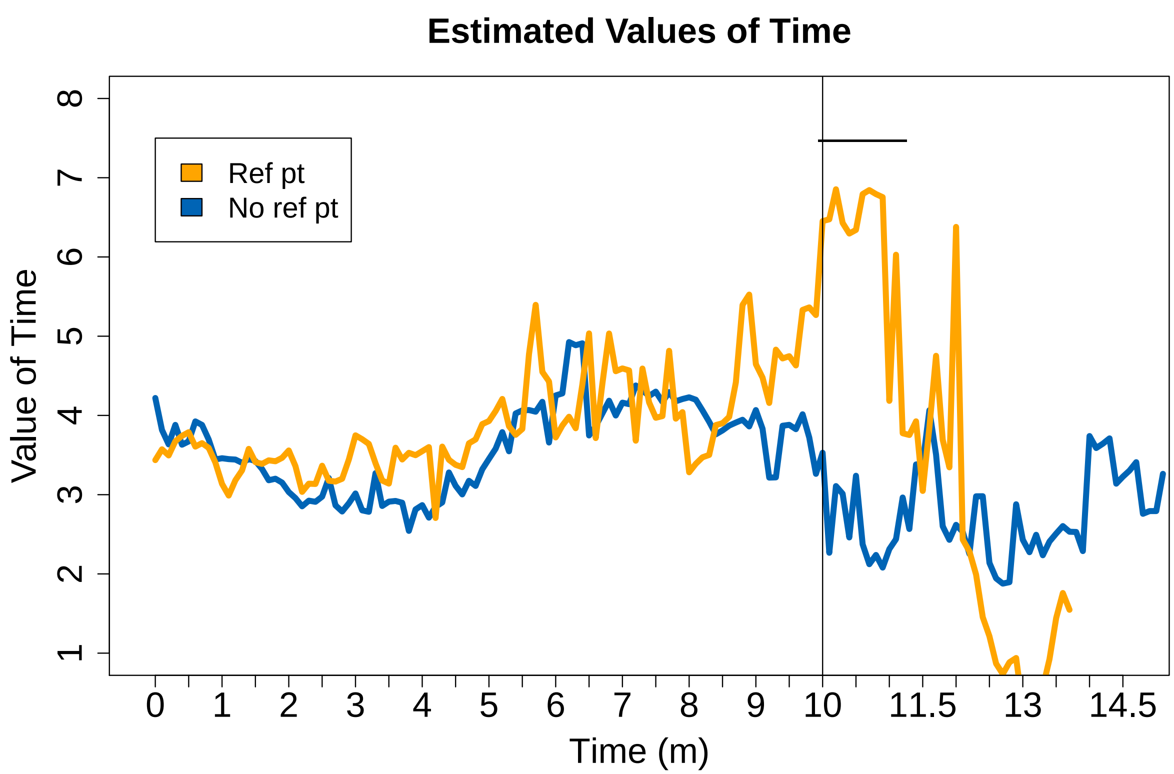

Figure 2.13: Moving average of estimated values of time

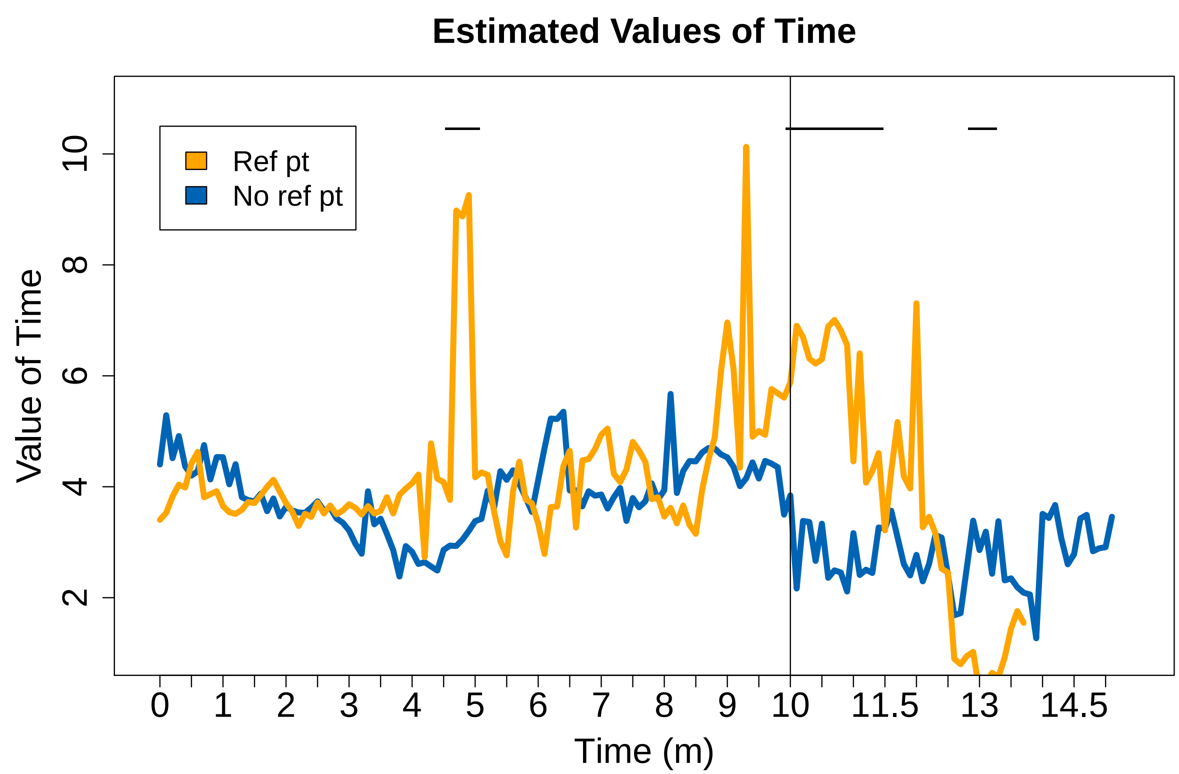

Figure 2.14: Higher resolution moving average of estimated values of time

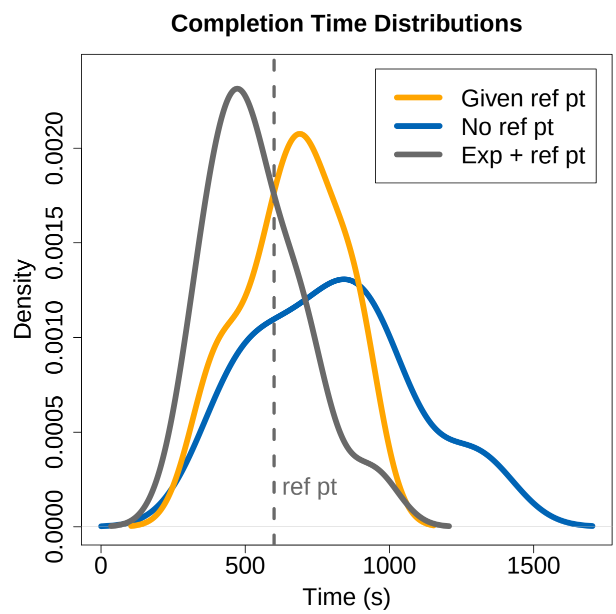

Figure 2.15: Kernel density estimate of completion time data with experienced participants

experienced participants, which peaks before the reference point is hit. Furthermore, the mean value of time shown in Figure 2.16, which accounts for changes in ability, is generally higher among those who are experienced, and peaks near the halfway point. These are not definitive markers of sophistication; they could stem from convexities in subjective cost due to fatigue or boredom from task repetition. Nonetheless, they are also consistent with forward-thinking agents, and suggest that reference dependence in time has the possibility to influence time allocation even outside of the loss regime.

The speed at which reference points adjust remains an unresolved issue in the empirical literature. To check whether experienced agents are influenced specifically by their previous behavior, I also plot in Figure 2.17 their values of time locked to their completion time in the first stage, along with 95% nonparametric bootstrap confidence intervals. Beyond a mild upward trend, no clear systematic reference point or benchmark effects are apparent before the first-stage completion time is reached. This suggests people may not be quickly adjusting their reference points based on direct experience. However, since little data is observed after the first-stage time passes, this remains inconclusive.

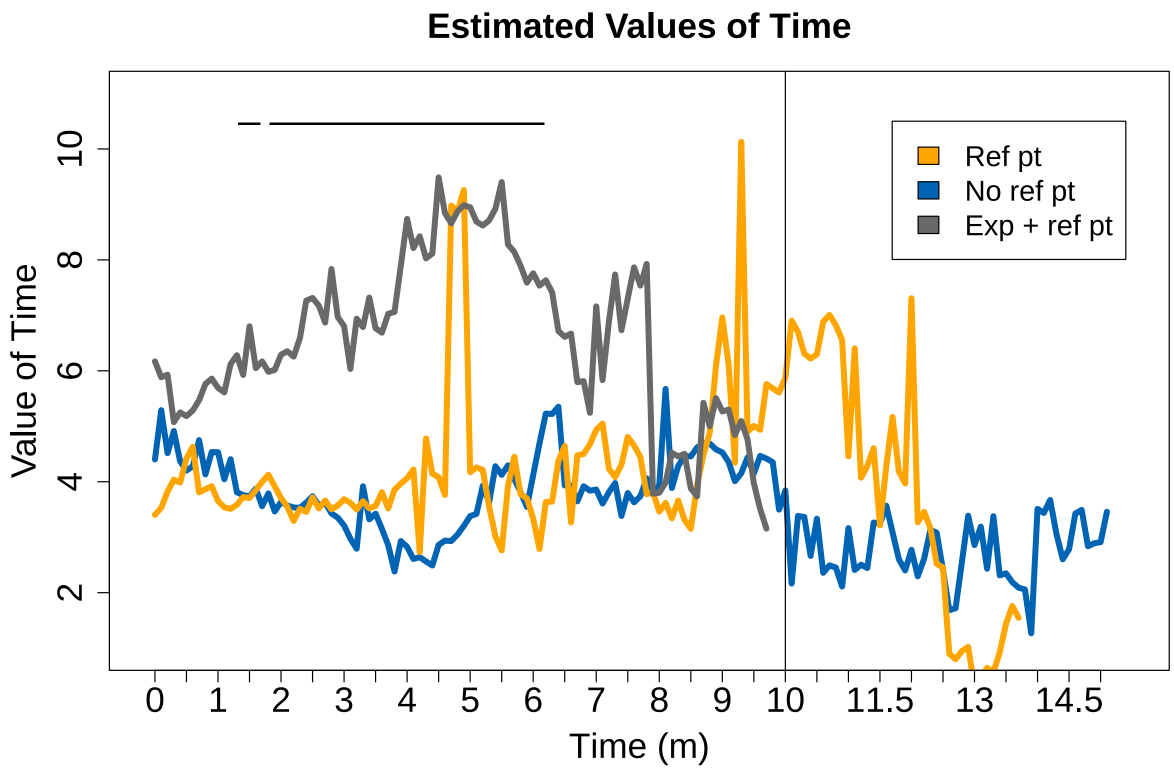

Figure 2.16: Moving average of estimated values of time with experienced participants

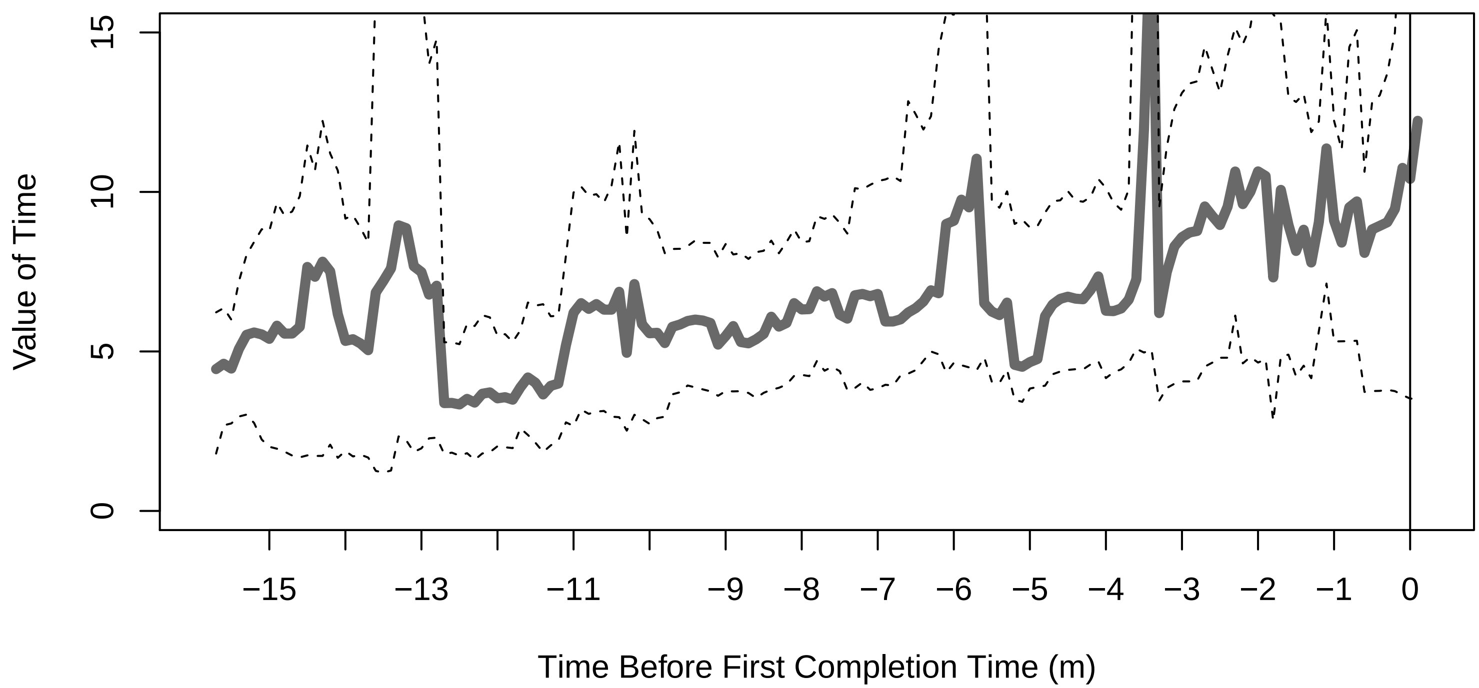

Figure 2.17: Moving average of estimated values of time with experienced participants relative to first stage completion time

2.5 Conclusion

People often have beliefs about how long tasks will take to complete and become discontent if their expectations are violated. According to theories of reference-dependent preferences (e.g. Bell, 1985; Loomes and Sugden, 1986; Gul, 1991; Kőszegi and Rabin, 2006), people will try to avoid falling too far behind these expectations even if it means forgoing monetary payoffs. I conducted a real-effort experiment to test the theory's predictions, finding support. Most directly, values of time more than doubled after participants exceeded the reference point based on external information, and work speed increased at the cost of reduced monetary earnings. Participant task completion times tended to cluster near the reference point. Those who fell behind were distinctly more dissatisfied, but not if they had prior experience in the task, suggesting that their reference points incorporated varied kinds of information. These findings strengthen the case for expectations-based reference dependence and the practical expansion of its domain to expectations in time.

My results apply to literatures on labor supply choices. I find evidence that the quantity and quality of labor supply are influenced by workers' beliefs independent of a link to pecuniary outcomes. Theories of reference-dependent preferences predict that among workers with flexible daily hours, stopping probabilities are related to income earned in a given day. Several studies show this pattern (Camerer et al., 1997; Chou, 2002; Fehr and Goette, 2007; Doran, 2014; Leah-Martin, 2015), but observed levels of variation in earnings are not fully explained by a fixed earnings target (Camerer et al., 1997; Farber, 2005; Farber, 2008; Farber, 2015). Reference points assumed to be based on expectations in both time and money may, however, be able to account for this by admitting flexibility in response to contextual variation (Kőszegi and Rabin, 2006; Crawford and Meng, 2011). My study provides a check on these assumptions, and the results help justify use of the theory. My quantitative estimates of the strength of time-based loss aversion also accord with the closest figures in the empirical literature based on an alternative structural model, further validating a reference dependence approach.

I also take a methodological step forward by exploiting a cognitively-inspired structure to identify parameters of economic interest. Variables not commonly used by economists can be powerful tools in understanding behavior, even variables that are simple to measure. My approach is in line with the recommendations of Rubinstein (2007) and Spiliopoulos and Ortmann (2014) to exploit response time data, but goes farther by interpreting those response times through the lens of a neuroeconomic model. The drift diffusion model implies a complex underlying structure where error enters via the belief updating process. This combined with economic motives generates observed patterns of choice. Building on ideas expressed in Clithero and Rangel (2013), I embed the drift diffusion model into a fuller theory, allowing us to estimate economic parameters which we would otherwise be unable to discern. Drawing on work from other disciplines in this manner is one way to create and justify theoretical structure. There are more opportunities in this rich vein slowly starting to be mined (Webb, 2013; Woodford, 2014). Neuroeconomic parameters are particularly "deep" in the sense of the Lucas critique; they describe the fundamental "technology" of decision making. These insights can be instrumentally, as well as intrinsically, useful.

Chapter 3

A STATISTICAL TEST FOR THE OPTIMALITY OF DELIBERATIVE TIME ALLOCATION

We are often tasked with choosing from multiple options. A subtle but essential part of the choice we make in selecting a job candidate or a consumer product is when to stop deliberating and pick an option. Such decisions involve an inverse relationship between speed and accuracy; we can make judgments that are fast but error-prone, or slow but high quality. Over the last half century, research in psychology and neuroscience has indicated a class of mathematical models of the deliberative process – diffusion models – that seem to well describe neural and behavioral data. However, while these models generate a speed-accuracy tradeoff, their applications are often agnostic as to how agents actually negotiate this tradeoff. One prominent hypothesis is that an agent's stopping criterion optimally balances the costs and benefits of spending time.

Do people optimally balance the costs and benefits of time spent and accuracy attained? When conditions change, does behavior change commensurately? The answers to these questions are important because they inform us about how we can best generalize, predict, and influence a person's behavior across various contexts. If people are behaving optimally, we can predict their behavior precisely using optimization models even when task demands move outside of their original confines. If, on the other hand, people are not behaving optimally, then alternative models will furnish better predictions, and there may be room for interventions to improve the efficiency of decision making. For instance, time limits could prevent people from spending excess time on chronic deliberation between similar courses of action where deliberating yields little return.

The answers are also challenging to determine, though, because costs and benefits are subjective. How one feels about spending time towards some end varies from person to person. Lacking ways to overcome this challenge, classic theoretical predictions derived from natural modeling approaches remain understudied. Past tests of optimality have been based on criteria that either neglect subjectivity or only coincide with expected utility theory under restricted circumstances. Nonetheless, to fully explore individual decision making, we must allow for individual preferences.

In this paper, I propose a flexible way to test expected utility maximization in stochastic time allocation settings. This test can be applied in a variety of scenarios spanning perceptual tasks or value-based decision making. In tandem, I conduct experiments in the perceptual domain to investigate optimality according to the most well-known of diffusion models, the drift diffusion model. The experiments include both simple decisions, in which the agent knows the difference in values between better and worse choices, as well as more complex decisions, in which the value difference is uncertain. The latter involves the first test of new theoretical results by Fudenberg, Strack, and Strzalecki (2015) characterizing the optimal decision rule for the uncertain-difference drift diffusion model.

I find that when the value difference is known, a substantial fraction of participants do not appear to respond optimally to changes in task difficulty. Furthermore, when the value difference is uncertain, although the optimal decision rule provides an improvement over the standard fixed threshold assumption, participants do not appear to be sensitive to changes in monetary incentives. Thus there is still significant room for improvement in understanding the process of deliberation.

3.1 Background

Among countless studies of decision making, a moderate number invoke mathematical portrayals of the deliberation process itself. A great deal of research measuring behavioral and, more recently, neural patterns has provided support for diffusion models (Ratcliff, 1978; Busemeyer and Townsend, 1993; LaBerge, 1962; Pike, 1966; Vickers, 1970; Usher and McClelland, 2001; Shadlen and Newsome, 1996; X.-J. Wang, 2002). These models describe decision making as one or more stochastic processes representing the evolution of confidence in one's answer as noisy information is processed. Once confidence reaches a particular threshold level, the agent stops processing information and commits to a choice. Originally developed in the mid 20th century, the drift diffusion model (DDM) is the oldest and most famous in this class of models (Stone, 1960; Laming, 1968; Ratcliff, 1978). Many pieces of behavioral evidence show that the DDM closely matches configurations of choice and response times in a variety of tasks ranging from perceptual discrimination (Ratcliff and Rouder, 2000; Ratcliff, 2002; P. L. Smith, Ratcliff, and Wolfgang, 2004) to recognition memory (Ratcliff, 1978; Starns and Ratcliff, 2014), and recent work extends it to value-based judgment (Krajbich, Armel, and Rangel, 2010; Milosavljevic et al., 2010; Krajbich and Rangel, 2011; Krajbich, Lu, et al., 2012). Moreover, direct measurements of neural activity reveal the implementation of evidence accumulation processes that fit the model's structure (Hanes and Schall, 1996; Shadlen and Newsome, 2001; Gold and Shadlen, 2002; Ratcliff, Cherian, and Segraves, 2003; P. L. Smith and Ratcliff, 2004).

Part of the DDM's original motivation was its formal analogy with efficient statistical algorithms. Importantly, the DDM was built as the continuous-time limit of the sequential probability ratio test (SPRT; A Wald, 1947) and is often theoretically characterized as inheriting its optimal stopping properties. For instance, it attains the speed-accuracy frontier (Abraham Wald and Wolfowitz, 1948; Arrow, Blackwell, and Girshick, 1949); that is, it achieves the highest accuracy for any given response time, and the quickest response time for any given accuracy level. Therefore questions of optimality bear upon the core identity of the model. We would like to understand how far the DDM's optimality extends in practice.

To be more specific, in a two-alternative forced choice, the key outcomes about which decision makers are thought to care are time spent and performance attained. The former is penalized due to objective or subjective costs of time, and the latter is rewarded by association with some payoff. In the special but commonly encountered case that occurs when the rewards from the correct and incorrect options are fixed, performance reduces to accuracy, the probability of choosing the correct option. There are then two natural criteria for optimality based on these central elements.

The first criterion is to minimize a weighted sum of error rate and decision time, which is known as the Bayes Risk (Abraham Wald and Wolfowitz, 1948):

$$\min_{x \in X} BR(x; y) = ER(x; y) + \psi DT(x; y),$$

where x and y are choice variables and fixed parameters, respectively, that determine the outcomes. The optimal choice depends on the free preference parameter ψ that determines how much importance is placed on time relative to performance, and which may include a sizeable subjective component that is not directly observable. Thus ψ is interpretable as the subjective flow cost of time. The Bayes Risk expression is a special case of expected utility maximization when the reward depends only on whether the response is correct; more generally, the decision maker is assumed to solve $\max_{x \in X} \mathbb{E}[reward(x; y) - cost \times time(x; y)]$. This is generally considered the most appropriate optimization criterion.

The second criterion is to maximize the ratio between accuracy and decision time, which is the Reward Rate (Gold and Shadlen, 2002):

$$\max_{x \in X} RR(x; y) = \frac{1 - ER(x; y)}{DT(x; y) + T_0 + D + ER(x; y) \times D_p},$$

where T_0 is the time required for sensory and motor processing, D is the time interval between a correct response and the following stimulus, and D_p is the additional time delay which penalizes an incorrect response on top of D. This expression is relatively more common in psychology and ecology due to its origins in reinforcement rate analysis, but remains rarely used outside of those fields. The optimal choice here does not require the specification of any free parameters.

If the environment is homogeneous in difficulty, then the SPRT (and DDM) maximizes any reward criterion based on error rate and decision time that is monotonically nonincreasing in decision time (Rafal Bogacz, E. Brown, et al., 2006). Thus the BR and RR optimal solutions happen to coincide perfectly. However, this is not the case in general. Because Reward Rate maximization is parameter-free, empirical tests of optimal choice have focused almost exclusively on this criterion (Simen et al., 2009; Rafal Bogacz, Hu, et al., 2010; Starns and Ratcliff, 2010; Starns and Ratcliff, 2012; Zacksenhouse, R Bogacz, and Holmes, 2010; Balci et al., 2011; Karşılar et al., 2014; Drugowitsch, Moreno-Bote, et al., 2012; Drugowitsch, DeAngelis, Klier, et al., 2014; Drugowitsch, DeAngelis, Angelaki, et al., 2015). Optimality in the Bayes Risk sense has thus been empirically neglected despite its theoretical importance.

I seek to rectify this imbalance by introducing a method to test expected utility maximization (and thus Bayes Risk optimality) in such stochastic decision settings. The test comprises a check for consistency of preferences across different conditions. The idea is that when changes occur in environmental parameters such as external incentives or problem difficulty, optimal behavior should change commensurately to reflect consistent underlying subjective preferences. Suppose, as in the following section, that we are interested in studying whether people can optimally adjust their deliberative behavior according to the difficulty of decision problems. Although problems of varying difficulties will lead to different ability levels and decision rules (which in the drift diffusion model are captured in the drift rate and confidence threshold) these differences should nonetheless reflect a consistent underlying preference. This consistency can be checked by estimating preference parameters across different sets of problems characterized by difficulty level, and testing whether preferences are the same even when difficulty varies.

That is, suppose one has data from two sets of problems, one with uniformly low difficulty and one with uniformly high difficulty. A person's ability is likely lower among the high difficulty problems. The optimal decision rule in each set will be based on one's ability and preference (for time versus accuracy). If we have reason to believe their preferences are the same across these sets (or at least hold some specific relation to each other), then the estimated decision rule in conjunction with the estimated ability in each set should imply the same preference across sets. This is the core of what is tested. While estimates of ability and decision rule will vary due to task difficulty, they should jointly indicate the same inferred preference if people are indeed optimally balancing the costs and benefits according to expected utility maximization.

Formally, as long as preferences can be identified using a type of likelihood estimation, consistency can be assessed using a likelihood ratio test where the restriction comes from the specified relation between preferences. In particular, if preferences are believed to be the same across multiple conditions, the restricted likelihood is based on equality between the preference parameters. Alternatively, weaker restrictions based on inequalities can be made if one only wants to assume ordinal relationships between conditions.

This method is in the spirit of economic tests of revealed preference; it attempts to rationalize behavior under some specified theory without making any claims about the intrinsic reasonableness of possible preferences beyond basic consistency. This constitutes a relatively minimal standard for rationality – consistency is necessary but not sufficient, and failure to reject the null hypothesis of consistency is not definitive evidence that the agent is behaving optimally.

The approach has natural advantages and limitations. The test relies on auxiliary assumptions about individual preferences. Because it is a consistency check, data is needed from multiple distinct but comparable conditions such that the analyst believes in some specific relationship between preferences across environments. Since preferences are allowed to be idiosyncratic, this imposes a meaningful constraint on when the test can be applied in practice. However, the test is flexible and can be applied in a general form with any model that makes precise predictions about the joint distribution of decision time and performance. The auxiliary assumptions can also be flexible; a variety of preference classes are permitted. This adaptability enables even more complicated models to be studied, including the Fudenberg, Strack, and Strzalecki (2015) DDM that I investigate in Section 3.3.

3.2 Variation in Difficulty

The choices we make often vary in their difficulty level. Some decisions are quick and obvious, while others are protracted and unclear. Can people optimally adjust their deliberative behavior according to the difficulty of decision problems? This ability is important for students taking their SATs or radiologists assessing the results of medical scans, all of whom must allocate time appropriately across easy and hard cases. I study this question in a common perceptual judgment paradigm, the random dot motion task, in which behavior and brain activity have been shown to fit the structure of the drift diffusion model.

Methods

In each trial of the random dot motion task, 100 white dots are displayed on a screen. As illustrated in Figure 3.1, a large fraction of them are "noise dots" moving in random directions (depicted as empty circles), while the remaining few are "signal dots" moving in a consistent direction (depicted as filled circles) — either all to the left or all to the right. The agent must determine in which direction the signal dots are moving. This is a common task in perceptual decision making experiments (e.g. Newsome, Britten, and Movshon, 1989; Britten et al., 1992) and is straightforward enough that similar versions have been administered to a range of nonhuman animals including rats and mice (Douglas et al., 2006), pigeons (Nguyen et al., 2004), and rhesus macaques (Kim and Shadlen, 1999).

Twenty-three human participants recruited from the Caltech SSEL population engaged in 100 trials of the standard random dot motion task. Difficulty was determined as usual by the number of dots moving in a consistent direction (the coherence). Trials were of 4 difficulty levels defined by coherences set at 10%, 12%, 14%, or 16% occurring uniformly at random throughout the block. Participants were paid a fixed amount of $0.05 for each correct answer and nothing for wrong answers, in addition to a $5 show-up fee. The timeline is depicted in Figure 3.2. Preceding each dot motion trial was a 1.5 second fixation cross. After each dot motion trial they received feedback as to which direction was correct in order to facilitate optimal behavior. The direction of coherent motion was determined with equal probability randomly across trials. The task was programmed and displayed using the Psychophysics Toolbox in MATLAB, with 5-pixel-width circular dots moving in a 960x960 pixel square aperture.

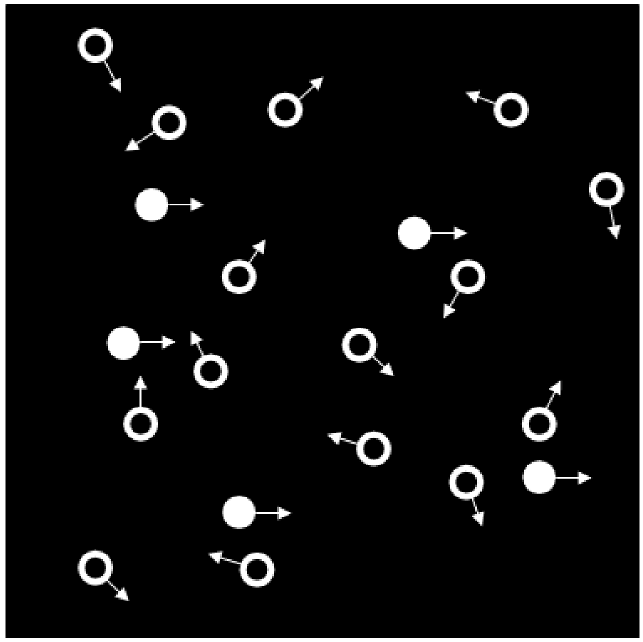

Figure 3.1: Schematic diagram of the random dot motion task stimulus

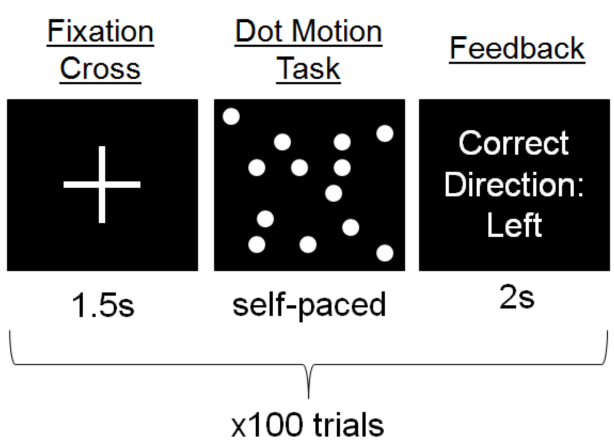

Figure 3.2: Timeline of experimental stimuli

Certain-Difference Drift Diffusion Model

The DDM models the process of choice as a stochastic accumulation of net evidence hitting an absorbing boundary. The model consists of four elements: the difference in evidence between the two alternatives at a point in time ($x(t)$), the drift rate that captures the agent's speed in integrating evidence (δ), the noise in the accumulation process (σ), and the threshold that stops the accumulation process ($\pm z$). Evidence for an option is integrated noisily in continuous time as represented by a Wiener process,

$$dx = \delta dt + \sigma dW,$$

that begins at $x(0) = 0$ in an unbiased decision, and stops as soon as it hits the decision threshold $\pm z$, at which time a choice is made according to whether $+z$ or $-z$ was hit. This threshold governs the speed-accuracy tradeoff. If it is set high, then the standard of evidence required to make a decision is stringent, and the resulting decision will be slow but accurate. If it is set low, then only weak evidence is needed, and the decision will be fast but inaccurate. This confidence threshold is chosen depending on one's preferences for time and accuracy; if optimal, it solves the Bayes risk minimization problem.

The properties of this model are well understood. Mathematically, its predicted outcomes arise as solutions to a first passage problem in stochastic processes — when does a Wiener process with drift first hit an absorbing boundary, and which does it hit? Analytical characterizations of accuracy and time contingent on its parameters are available due to its tractability (Rafal Bogacz, E. Brown, et al., 2006). The error rate is given as

$$ER = \frac{1}{1 + e^{2\delta z/\sigma^2}},$$

and the mean decision time is given as

$$DT = \frac{z}{\delta} \tanh\left(\frac{\delta z}{\sigma^2}\right).$$

The optimal decision threshold z^* is thus

$$z^* = \arg\min_{z \in \mathbb{R}_{++}} BR(z; \sigma, \delta, \psi) = \arg\min_{z \in \mathbb{R}_{++}} ER(z) + \psi DT(z),$$

and is the solution to

$$\frac{1}{\psi}\frac{2\delta^2}{\sigma^2} - \frac{4\delta z^*}{\sigma^2} + e^{-(2\delta z^*/\sigma^2)} - e^{2\delta z^*/\sigma^2} = 0.$$

Although this does not produce a closed form solution, it is known to uniquely determine z since the equation can be written as an equality between increasing and decreasing functions (Rafal Bogacz, E. Brown, et al., 2006).

Test Procedure

The optimality test centers on checking for consistency of individual preference parameters across difficulty levels (i.e. coherences). To do so, the DDM parameters $(z_i, \delta_i, \sigma_i)_{i\in\{10,12,14,16\}}$ must be estimated in each condition via one of several known psychometric methods. From each of these estimates, the corresponding preference parameters $(\psi_{10}, \psi_{12}, \psi_{14}, \psi_{16})$ can be recovered supposing the optimal decision threshold balances the subjective costs and benefits of time and accuracy which are summarized by the preference parameter: $z = \arg\min BR = \arg\min a(z) + \psi t(z)$. Finally, these estimates of ψ can be checked for consistency, $\psi_{10} = \psi_{12} = \psi_{14} = \psi_{16}$. Intuitively, this is checking that the subjective cost of time implied by behavior is the same across levels of difficulty.

I propose and implement a general hypothesis testing approach that relies on a combination of highly flexible tools and can thus be applied to a wide range of models. This approach is applicable as long as the model makes specific predictions about the joint distribution of time expenditure and accuracy, even if that distribution is analytically intractable.

1. Construct and maximize a likelihood-type function based on distributions $f(t, \theta)$ simulated with Monte Carlo methods to estimate the model parameters θ for each condition:

$$\log \mathcal{L}(\theta) = \sum \log f(t, \theta).$$

2. Test for differences in the cost of time parameter ψ across conditions using a likelihood ratio test in which the unrestricted likelihood, $\mathcal{L}_{ur}$, allows ψ to be free in each condition while the restricted likelihood, $\mathcal{L}_r$, forces ψ to be equal across conditions (with degrees of freedom df_{ur} and df_r equal to the number of model parameters under each assumption):

$$2(\log \mathcal{L}_{ur} - \log \mathcal{L}_r) \sim \chi^2_{df_{ur}-df_r}.$$

To implement this procedure and characterize the full likelihood function, first I simulate time-accuracy distributions contingent on values of the model parameters $\theta = (\psi, \delta, \sigma)$. These parameter values are taken from a 30x20 grid with ψ varied from 0.001 to 0.030 by increments of 0.001 and δ varied from 0.005 to 0.1 by increments of 0.005. The properties of the model are known to depend only on the relative values of the parameters, so as is customary, σ is taken to be 0.1. In order to generate each distribution, as described below I simulate a first passage problem using a random walk approximation (Tuerlinckx et al., 2001). Wiener processes are often described as the continuous time limit of a discrete random walk (e.g. Feller, 1968; Cox and Miller, 1977), which lends itself naturally to simulation. A random walk on a discrete state space with tiny displacements and tiny time intervals can be used to simulate a Wiener process with drift δ and variance σ^2 as follows. In every time interval τ, the state increases by a fixed displacement $\Delta = \sigma\sqrt{\tau}$ with probability p and decreases by Δ with probability $1 - p$, where

$$p = \frac{1}{2}\left(1 + \frac{\delta\sqrt{\tau}}{\sigma}\right).$$

As τ converges to zero, this approximation increasingly resembles the continuous time stochastic differential equation, but the algorithm's computational burden also grows. I set $\tau = 10ms$ to produce the most accurate representation possible in a reasonable amount of computing time. The time-accuracy distribution is then computed based on how many time steps the random walks take to first cross the decision threshold $\pm z^*$ that stems from the grid point values ψ and δ (and σ) and which threshold of $+z^*$ or $-z^*$ is crossed first. This produces a simulated distribution $f(t)$ for each grid point θ. I generate 10,000 random walks per parameter grid point to construct these distributions. This process is depicted in Figure 3.3; sample random walks are shown hitting either the correct (top) or incorrect (bottom) nonlinear decision threshold at various time points, which gives rise to a simulated time-accuracy distribution.

The simulated time-accuracy distributions for candidate model parameters are then compared with the data to create a likelihood-type function. I use quantile maximum probability estimation (QMPE; Heathcote, S. Brown, and Mewhort (2002), Heathcote, S. Brown, and Cousineau (2004), and S. Brown and Heathcote (2003)), which is a more robust alternative to standard maximum likelihood estimation. QMPE groups data into bins based on response time quantiles, $q_{j=0,\dots,m}$, so that the predicted probability of a data point is given by the proportion of samples falling into its bin, $\pi_j(\theta) \equiv \int_{q_{j-1}}^{q_j} f(t,\theta)dt$. To achieve a reasonable balance between robustness and efficiency, I construct bins from deciles ($m = 10$). The function to be maximized

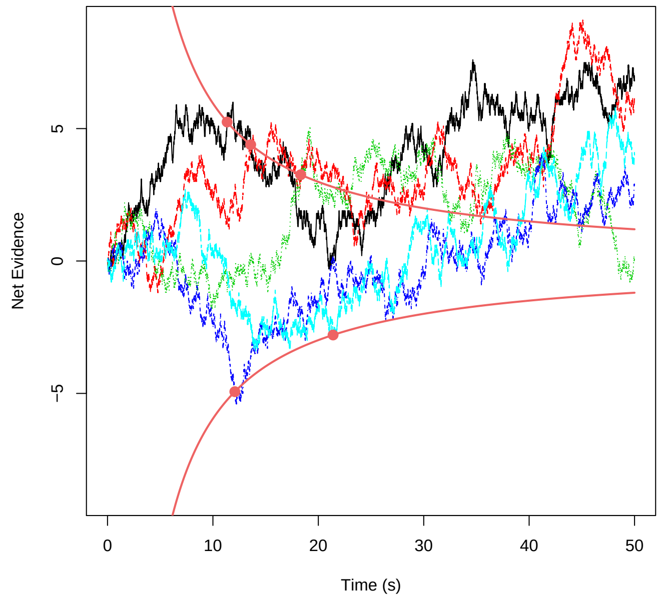

Figure 3.3: Illustration of random walk simulations

then takes the form of the multinomial log likelihood:

$$\log \mathcal{L} = \sum_{condition} \sum_{j=1}^{m} (N_{j,\text{Correct}} \log \pi_{j,\text{Correct}} + N_{j,\text{Error}} \log \pi_{j,\text{Error}}).$$

This yields parameter estimates $\hat{\theta}$ that are consistent and asymptotically normally distributed (Menéndez et al., 2001; S. Brown and Heathcote, 2003). Figure 3.4 depicts a likelihood function created using this procedure.

QMPE has been shown to exhibit superior performance compared to standard ML approaches when applied to response time distributions (Heathcote, S. Brown, and Mewhort, 2002; Heathcote, S. Brown, and Cousineau, 2004), and has attained great popularity in the perceptual decision making literature due to its merits. Because the method is grounded in order statistics, it is relatively robust to outliers and contamination; one can easily see that making extremal data points even more extreme will not affect the estimate. The binning process will furthermore help smooth over bumps in the simulated distributions. However, superior performance may also be linked to hidden theoretical advantages of the estimator.

QMPE turns out to be a branch of maximum product of spacings estimation (MPS; Cheng and Amin, 1983; Ranneby, 1984). MPS is predicated on the fact that the percentiles of any continuous distribution are themselves uniformly distributed, based on the probability integral transform. Whereas MLE tries to directly maximize the fit of the predicted distribution to the data, MPS tries to maximize the fit of the uniform distribution to the data percentiles. The estimator maximizes the log of gaps between values of the distribution function at adjacent data points; that is, if an ordered sample $x_{(1)}, \ldots, x_{(n)}$ comes from a cumulative distribution function $F(x; \theta_0)$, spacings are defined as $D_i(\theta) = F(x_{(i)}; \theta) - F(x_{(i-1)}; \theta)$, and the MPS estimator is $\hat{\theta} = \arg\max_{\theta \in \Theta} \sum_i \ln D_i(\theta)$. This can be derived as an alternative estimator of the Kullback-Leibler divergence (Ekström, 2008). While MPS is generally asymptotically equivalent to MLE (Anatolyev and Kosenok, 2005) and shares its properties when the MLE exists, MPS can remain consistent and efficient even when MLE fails. In practice, Monte Carlo simulations provide evidence of greater efficiency and lower bias in QMP estimates compared to MLE among settings that characterize perceptual decision making (Heathcote, S. Brown, and Mewhort, 2002; Heathcote, S. Brown, and Cousineau, 2004).

Finally, based on the simulated likelihood function, a likelihood ratio test (LRT) can assess whether the preference parameter ψ is consistent across conditions. The restricted likelihood accordingly assumes the null hypothesis $H_0 : \psi_{10} = \psi_{12} = \psi_{14} = \psi_{16}$, and the unrestricted likelihood assumes the alternative hypothesis $H_1 : \psi_{10} \neq \psi_{12} \neq \psi_{14} \neq \psi_{16}$. The test has degrees of freedom $df_{ur} - df_r$ based on the difference in the number of unrestricted and restricted model parameters. The former allows a different preference parameter ψ for each condition, whereas the latter requires the same single parameter in each condition, so the difference is equal the number of conditions less one (in this case, $df = 3$). The test statistic $2(\log \mathcal{L}_{ur} - \log \mathcal{L}_r)$ is then distributed $\chi^2_{df_{ur} - df_r}$. LRTs are quite general and accommodate multiple parameters and multiple conditions. Due to the generality of Wilks's (1938) theorem which implies that likelihood ratio test statistics are distributed as χ^2, most likelihood-type estimation techniques should lend themselves well to this test. MSP shares the asymptotic properties of MLE, indicating that an LRT remains suitable (Ekström, 2013).

Beyond this, due to ceiling (100%) accuracy among many of the participants in several of the conditions, I employ an imputation scheme in which an extra data point is appended. This imputed trial comprises an erroneous choice that occurs in the lowest response time quantile. In line with the speed-accuracy tradeoff, the idea is that a participant would make a mistake if they had responded even quicker than their quickest observed time. This imputation was applied to the 14 of 23 participants who exhibited ceiling accuracy in at least one condition due to a combination of low difficulty and few data points per difficulty level. Regardless, the test rejects optimality for approximately the same proportion of individuals even when considering only the subset who required no imputation.

Results

Figure 3.4 shows for each difficulty condition the likelihood functions of a sample individual for whom optimality was rejected. White grid points denote the best fitting parameter configuration, and surrounding yellow areas represent the χ^2_2 95% confidence region. The drift rate parameter which captures one's information accumulation ability can be seen to rise as the difficulty level drops, as would be expected if the experimental treatment has an effect. The test, though, does not require commitment to relationships between non-preference parameters and therefore does not strictly depend on this. However, the preference parameter representing the subjective cost of time varies with difficulty, reflecting the kind of inconsistency the test is intended to pick out.

This test was implemented for each individual in the sample. The histogram of resulting test p-values is displayed in Figure 3.5. Optimality was rejected for 6 out of 23 people at the 5% significance level. We can diagnostically examine the estimated parameters across conditions. Figure 3.6A reveals that overall the drift rates are indeed increasing as the task becomes easier, while Figure 3.6B reveals greater variation in estimated preferences among those the test rejected. This indicates that people for whom optimality is not rejected are also affected by the treatment to some degree. Thus the test is not merely detecting any kind of change due to the treatment but is selectively picking out violations of preference consistency. Nonetheless, a sizeable portion of the preference variation does appear to be related to the dramatic effect of the hardest difficulty level. This suggests that stronger treatments inducing greater differences in difficulty could reveal further inconsistency, highlighting the point that the test is a relatively minimal criterion and individuals who do not fail it may yet be found suboptimal under other circumstances.

As mentioned earlier, the test requires the relationship between preference parameters across conditions to be specified. Consistency in the present case was taken to mean that preferences should be identical across difficulty levels; that is, $\psi_{\text{higher difficulty}} = \psi_{\text{lower difficulty}}$. However, we may believe that time spent on harder

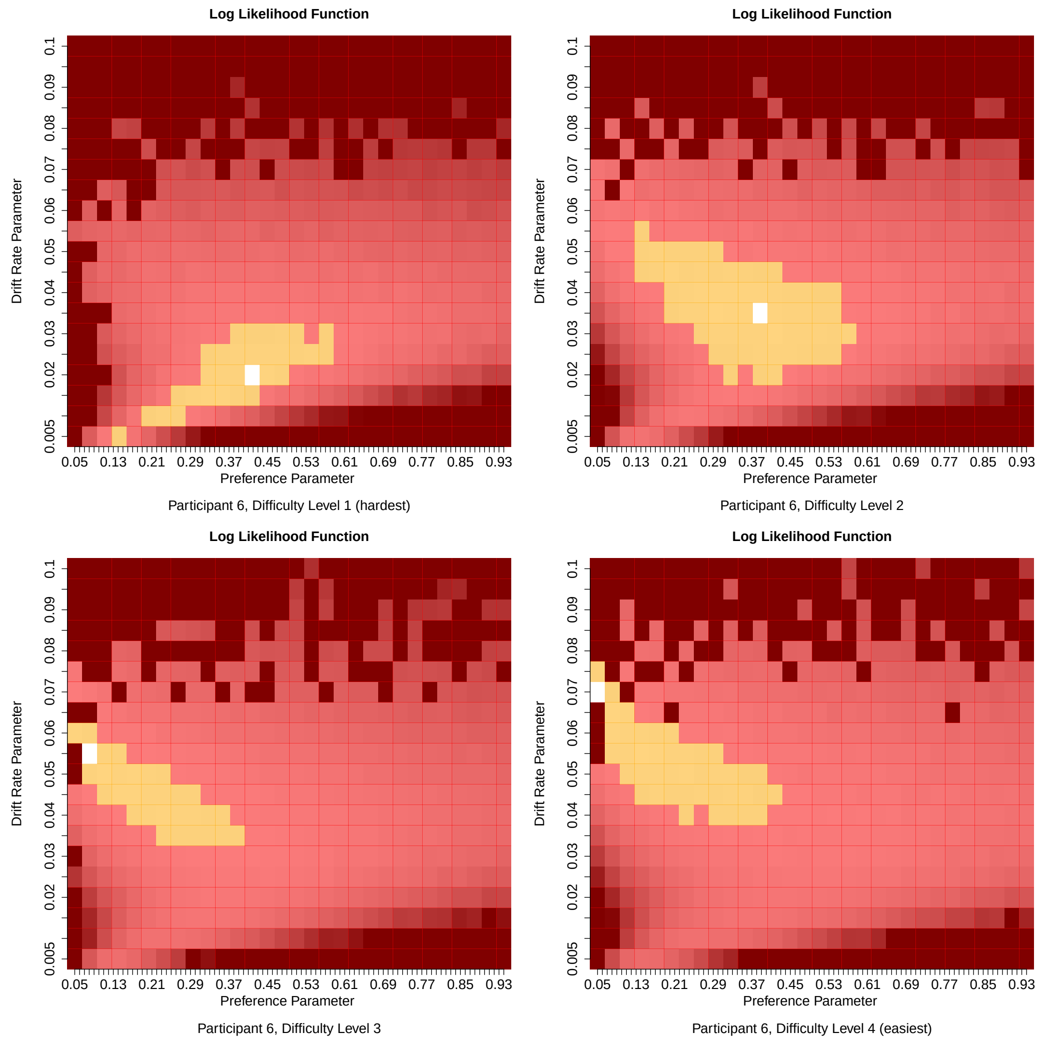

Figure 3.4: Example of simulated likelihood functions

problems could feel subjectively worse than the same amount of time spent on easy problems. This would imply that preferences can change and still reflect an underlying consistency, constituting an alternative definition of consistency which could rationalize the behavior of some individuals who were previously judged to be inconsistent. In fact, the test can be adapted to this alternative criterion by relaxing the equality restriction while maintaining the same general approach. Rather than assuming $\psi_{\text{higher difficulty}} = \psi_{\text{lower difficulty}}$, I enforce an order-restricted hypothesis, $\psi_{\text{higher difficulty}} \geq \psi_{\text{lower difficulty}}$.[1] Using a conservative bound, the test still rejects

[1]Under such conditions, theoretical results imply that the LRT statistic is no longer distributed as χ^2. Instead the test statistic is usually distributed as $\bar{\chi}^2$, which is a mixture of χ^2 distributions with varying degrees of freedom (e.g. Robertson, 1978; Robertson, Wright, and Dykstra, 1988). Intuitively, the $\bar{\chi}^2$ distribution formalizes what can be thought of as a χ^2 distribution with non-

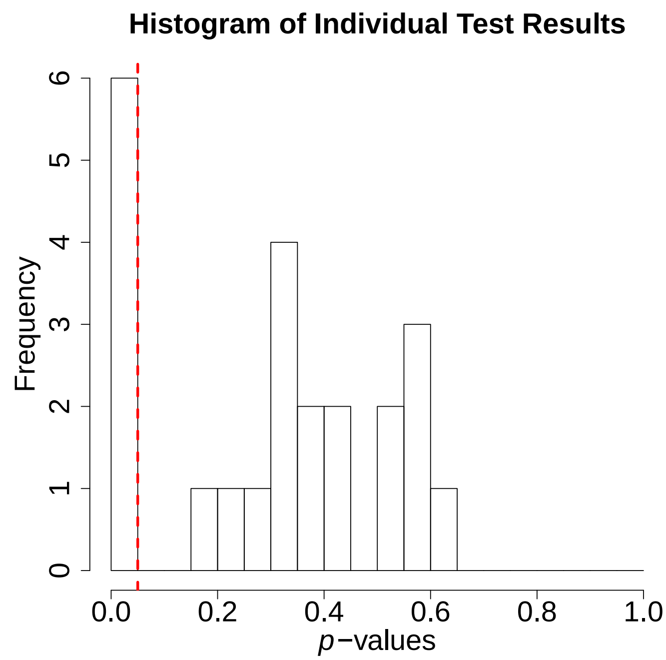

Figure 3.5: Histogram of individual test results

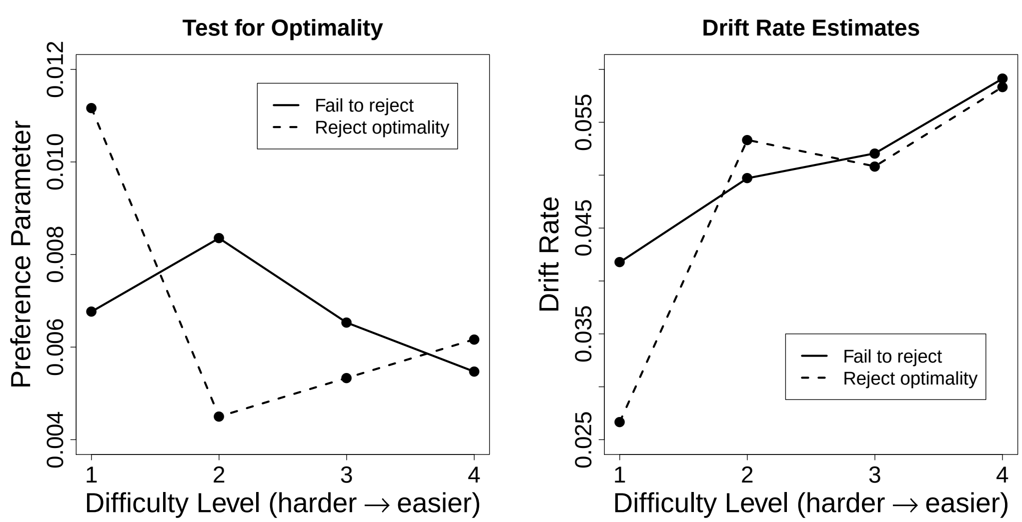

Figure 3.6: Histogram of estimated parameters by group

optimality for at least two individuals.

3.3 Variation in Incentives

A central tenet of economic reasoning is that incentives – usually monetary in nature – matter. Do people respond optimally to changes in monetary incentives in a deliberative context? I move beyond the usual paradigm by further studying optimal time allocation in an environment where the difference in payoffs between alternatives is uncertain. I do so using a modified version of the random dot motion task enabling me to conduct the first test of Fudenberg, Strack, and Strzalecki's (2015) theory, a new and precise characterization of collapsing thresholds derived as the solution to an optimization problem. The task forms an intermediate step between perceptual and value-based decision making in which I can test the theory under precise and fitting circumstances.

In Section 3.2, the difference in values between a correct and an incorrect answer was assumed to be fixed. Under this assumption, expected reward boils down to accuracy. However, in many situations the difference in values may be variable and uncertain. Such uncertainty often occurs in value-based decision making, to which the DDM is increasingly being applied (Krajbich, Armel, and Rangel, 2010; Milosavljevic et al., 2010; Krajbich and Rangel, 2011; Krajbich, Lu, et al., 2012). For instance, an employer deciding between two potential employees is unsure not only about which of them is better, but also by how much. One candidate might substantially increase profits while the other ruins a project, in which case lengthy deliberation is crucial – or both might be about equally effective, in which case lengthy deliberation will not produce much additional value.

In the value-based DDM, the accumulation rate is based on the difference in values between options. If one option is much better, then confidence tends to rise quickly. Thus the confidence trajectory carries information about the value gap in addition to option rank. Spending a long time deliberating without coming to a conclusion accordingly implies that the options are similar in value. The agent should then curtail their deliberation time. This means the optimal decision threshold will no

integer degrees of freedom. Consider that restricting two parameters to be equal to each other removes a single degree of freedom; it follows that restricting one parameter to be weakly greater than the other is less stringent and should in some sense remove only a fractional degree of freedom. In practice, the mixture weights for each χ^2 component of the $\bar{\chi}^2$ distribution are difficult to calculate, nor do the most relevant results currently exist for order-restricted inference with MPS. However, the $\bar{\chi}^2$ critical value will be bounded by the χ^2 critical value with the equality-restricted degrees of freedom.

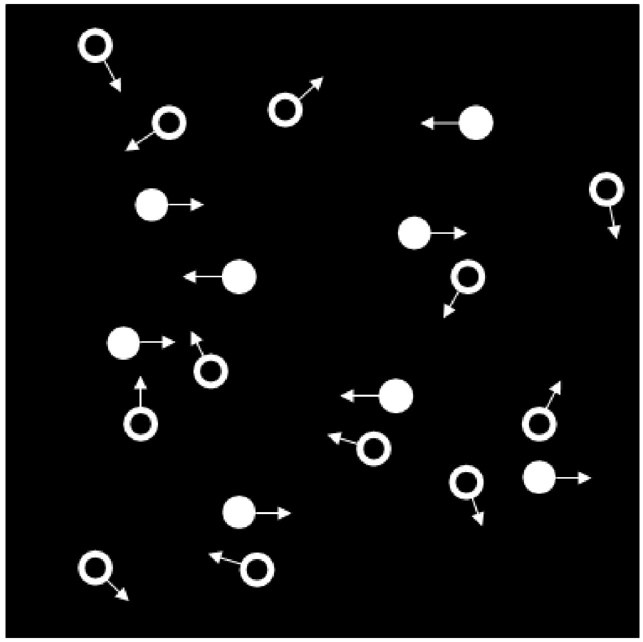

Figure 3.7: Schematic diagram of the bi-directional random dot motion task stimulus

longer be fixed; it will instead collapse over time. But how fast should it collapse? Fudenberg, Strack, and Strzalecki (2015) develop analytical results characterizing the optimal threshold contingent on parameters including the cost of time parameter. My proposed test is flexible enough to accommodate even this more complex theory.

Methods

Rather than the usual dot motion stimulus consisting of two groups of dots in each trial – signal dots moving consistently left or right, and noise dots moving randomly – the modified dot motion stimulus comprises three groups of dots – signal dots moving consistently left, signal dots moving consistently right, and noise dots moving randomly. This is displayed in Figure 3.7. Participants can choose left or right, and the wage earned is proportional to the number of dots moving in the chosen direction (the coherence). That is, if 30 dots are moving left and 20 dots are moving right, picking left might earn \$0.12 while picking right might earn \$0.08. Lam (2014) studies this task and finds that the difference between left- and right-coherences is indeed what drives the accumulation process. In this way, the drift rate is tied to value as assumed by the value-based DDM.

Twenty-four participants (less 1 excluded due to computer error) recruited through the Caltech SSEL engaged in two blocks of 100 trials each. On top of a \$10 show-up fee, they earned points equal to the number of dots moving in the direction they chose. The pay rate in the first block was set at \$0.02 per 10 points, and was doubled in the second block to \$0.04 per 10 points. In line with the theoretical assumptions (laid out below), the numbers of signal dots moving in each direction were drawn i.i.d. from a discretization of the normal distribution $\mathcal{N}(25, 7)$ of which participants were explicitly informed. In order to facilitate optimal behavior, they received feedback on the correct direction following each trial.

Uncertain-Difference Drift Diffusion Model

The agent is faced with two options, $i = l, r$, which have unknown values, $(v^l, v^r) \in R^2$. She holds a prior belief about these values, $\mu_0 \in \Delta(R^2)$, and observes a signal $(Z_t^i)_{r \in R_+}$ which, as in the DDM, drifts according to the Wiener process:

$$dZ_t^i = v^i dt + \alpha dB_t^i.$$

She continuously updates her belief about the values, holding a posterior mean for v^i denoted $X_t^i = E[v^i | \{Z_s^i\}_{0 \leq s < t}]$ that is conditioned on the signal trajectory up to each point in time.

The agent must decide both which option to select and when to stop deliberating. She chooses the option with the highest posterior expected value when she stops, and also chooses a stopping time t. A flow cost $\psi > 0$ is incurred for time spent before an option is selected. Thus she confronts the Wald optimality problem

$$\max_{\tau \in T} \mathbb{E}\left[\max_{i=l,r} X_\tau^i - \psi \tau\right].$$

A sufficient statistic for this decision is the difference in signal values,

$$Z_t \equiv Z_t^l - Z_t^r = (v^l - v^r)t + \sigma B_t,$$

where $\sigma = \alpha\sqrt{2}$ and $B_t = \frac{1}{\sqrt{2}}(B_t^l - B_t^2)$ is a Brownian Motion.

Both the usual certain-difference DDM and uncertain-difference DDM can be modeled in this framework; the key amendment lies in the prior belief. The certain-difference model confines the state space to two possibilities, in both of which one option is better than the other by a fixed positive amount. The agent knows she is not indifferent even before coming to a conclusion. If she has spent a long time and Z_t is still close to 0, she essentially faces the same problem she started with, and is thus as willing to continue with deliberation as she was at the outset. The current value of the process Z_t is a sufficient statistic for stopping (i.e. the past trajectory carries no additional useful information), and so the stopping threshold does not change over time (Shiryaev, 1969).

The uncertain-difference model instead assumes a Gaussian prior on option value, $v^i \sim \mathcal{N}(X_0, \sigma_0^2)$. Then spending a long time without a conclusion implies the agent is probably nearly indifferent between options, and should therefore cut short her deliberation and decide quickly. Thus the stopping threshold decreases over time. Fudenberg, Strack, and Strzalecki (2015) characterize the optimal stopping rule and derive a functional form that approximates the solution well according to their simulations:

$$z^*(t) = \frac{1}{2\psi(\sigma_0^{-2} + 2\sigma^{-2}t)}.$$

This function declines asymptotically to zero, meaning that eventually the agent chooses almost at random. It is hyperbolic as a function of time, and shaped by the prior variance σ_0, the signal noise σ, and, importantly for present purposes, the cost of time ψ. Standard static threshold models formally encode prior information primarily by biasing the starting point of the accumulation process. In contrast, prior information actually affects the shape of the threshold in the uncertain-difference model even though the accumulation process remains unbiased, as can be seen from the role of σ_0. For example, if the agent has a high-variance prior over value (i.e. σ_0 is large), she will require more evidence to make a selection in order to exploit the high-value trials which she believes will occur frequently. Notice further that this model does not nest constant thresholds as a special case, in contrast to the more flexible collapsing threshold models that are usually applied (e.g. Hawkins et al., 2015).

Test Procedure

The test proceeds in a roughly similar fashion as before, though experimental conditions are determined by payoff regimes. The drift rate, denoted δ, varies by trial and is proportional to the difference in the number of dots moving in each direction in a given trial, which is known to the experimenter. I construct the likelihood function by grid search to ensure global maxima are found. I simulate the time-accuracy distributions for candidate model parameters ψ and σ (which are allowed to vary across blocks) taking into account the assumed parameters δ and σ_0 set experimentally. The 90x90 grid varies ψ from 0.05 to 0.94 by 0.01, and σ from 10 to 188 by 2. I generate 10,000 random walks per parameter grid point (including the variation in δ; thus approximately 400,000 random walks per (ψ, σ) combination)

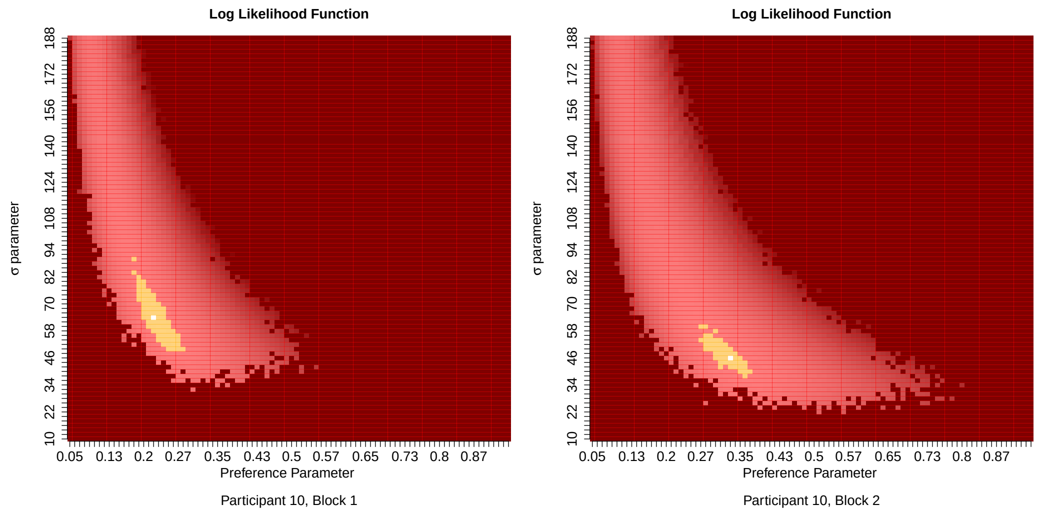

Figure 3.8: Illustration of simulated likelihood function for uncertain-difference model

to construct these distributions. These procedures are extremely computationally intensive; to run the simulations and construct likelihood functions for all 23 participants required on the order of 5,000 core-hours of computing time (4,000 and 1,000 core-hours respectively). This analysis was conducted in R, taking advantage of parallel processing with Amazon EC2. Sample likelihood functions for one individual are shown in Figure 3.8. As before, the white dot represents the best fitting parameter set and the yellow area represents the 95% confidence region.

Results

As a check, increased coherence differences did indeed translate closely into improved performance in a steady fashion, shown in Figure 3.9. (The transparency level of the plotted points represents the amount of data in each bin.)

The preference parameter represents the cost of time relative to reward. When the reward doubles, this parameter should then be cut in half. Figure 3.10 depicts how the estimated preference parameter actually changes across payoff regimes. Only a single person responds in line with this prediction represented by the null hypothesis of $\psi_{\$} = 2\psi_{\$\$}$. Many people exhibit no response to the change in payoffs. Intuitively this appears to be a kind of "stickiness"; these individuals are spending similar amounts of time regardless of the environmental conditions. Several people exhibit increased subjective costs. This may be due to fatigue caused by the task. This highlights the centrality of auxiliary assumptions about preferences; the test requires

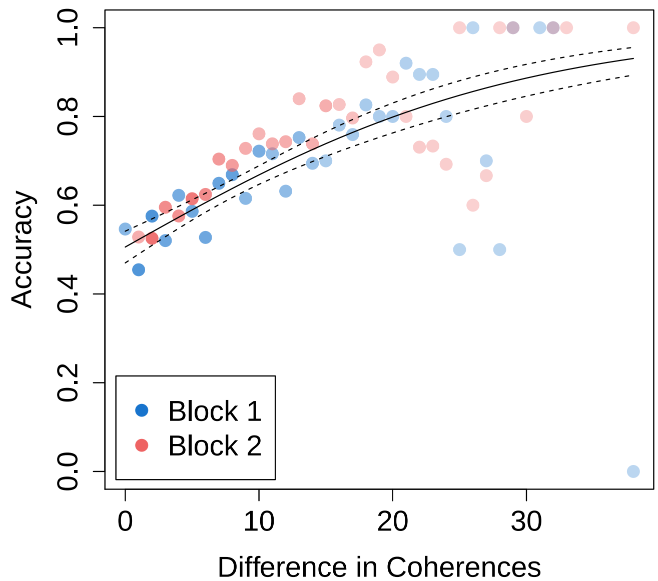

Figure 3.9: Relationship between coherence and accuracy

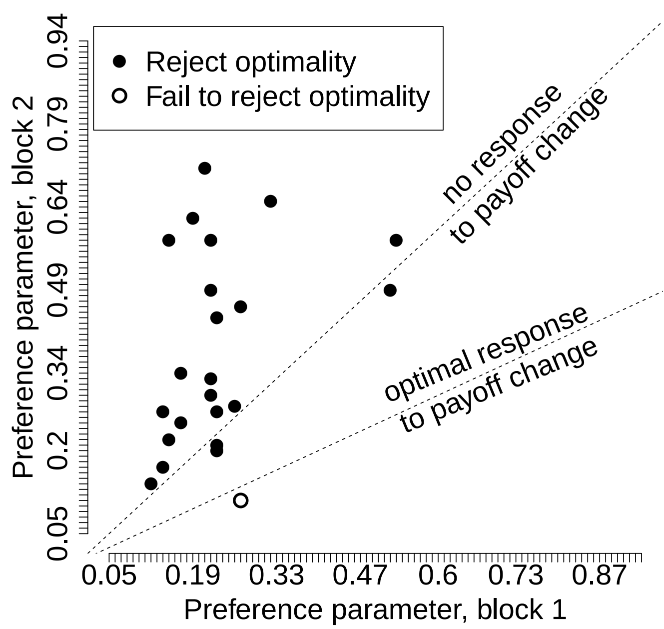

Figure 3.10: Scatterplot of individual preference consistency

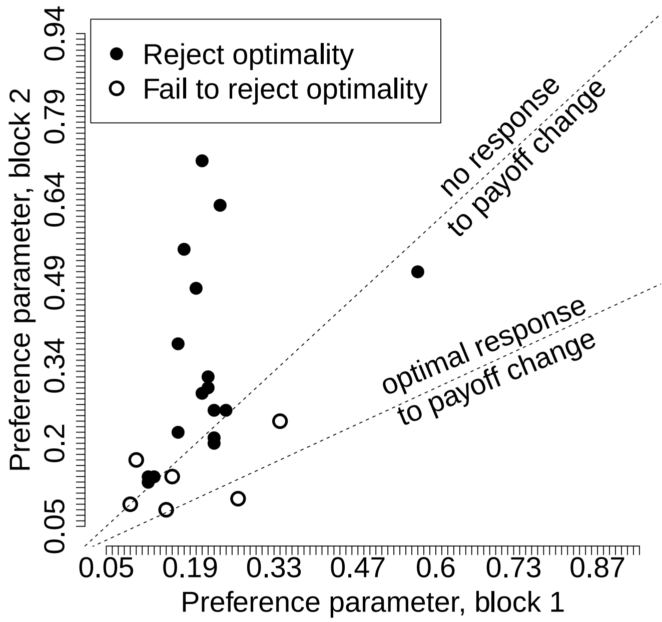

Figure 3.11: Scatterplot of individual preference consistency allowing cost of time to increase

some specific commitment to be made about how preferences should change across conditions, and any factors one believes to be important need to be encoded in the model and the test. Thus alternative specifications may be able to rationalize the behavior of more people.

To account for the effect of fatigue on preferences, I estimate a more complex model allowing the cost function to increase. The previous model assumed that $\psi = \psi_\$ \times (1 + \mathbb{I}(\text{2nd block})\psi_\Delta)$, and the test assessed whether $\psi_\Delta = 0$. Now I consider ψ to also increase linearly as trials pass so that $\psi = \psi_\$ \times (1 + \mathbb{I}(\text{2nd block})\psi_\Delta) + \psi_\tau \tau$, where τ is the trial number and ψ_τ is an additional parameter estimated on a grid from 0 to 2e−3 by increments of 2e−5. (Since ψ represents a marginal cost, allowing it to increase linearly in trial number is similar to having a quadratic total cost of time.) The test still assesses whether the change due specifically to the experimental block, ψ_Δ is 0. The results are shown in Figure 3.11. The test now rejects optimality for 17 individuals, 74% of the sample. Although this is fewer people compared to the simpler preference specification, the number remains high.

Relative Fit

Because this is the first test of the Fudenberg, Strack, and Strzalecki (2015) uncertain-difference DDM, I further study how well the model fits behavior relative to a standard static threshold model. Whether constant bounds or time-varying bounds should be preferred is a long-standing and contentious issue. Although many have satisfactorily used fixed threshold models (e.g. Bode et al., 2012; J. W. Brown et al., 2008; Ding and Gold, 2010; Ding and Gold, 2012; Forstmann, Dutilh, et al., 2008; Forstmann, Anwander, et al., 2010; O'Connell, Dockree, and Kelly, 2012; Ramakrishnan and Murthy, 2013; Ramakrishnan, Sureshbabu, and Murthy, 2012; Ratcliff, Philiastides, and Sajda, 2009; Salinas and Stanford, 2013; Schall, 2003; Schurger, Sitt, and Dehaene, 2012; P. L. Smith and McKenzie, 2011; Usher and McClelland, 2001; X.-J. Wang, 2002; Wong and X.-J. Wang, 2006), a substantial number have argued that collapsing thresholds fit data more closely (Sanders and Ter Linden, 1967; Paolo Viviani, 1979a; Paolo Viviani, 1979b; P Viviani and Terzuolo, 1972; Ditterich, 2006a; Ditterich, 2006b; Churchland, Kiani, and Shadlen, 2008; Cisek, Puskas, and El-Murr, 2009; Rao, 2010; Hanks et al., 2011; Bowman, Kording, and Gottfried, 2012; Thura, Beauregard-Racine, et al., 2012; Thura and Cisek, 2014; Zhang et al., 2014). Hawkins et al. (2015) carried out the most comprehensive meta-analysis to date and found evidence primarily in favor of the fixed threshold DDM as compared to the collapsing threshold DDM or urgency gating model. The follow-up study of Voskuilen, Ratcliff, and P. L. Smith (2016) concurred with this conclusion.

Time-varying thresholds in diffusion models are often defended on the basis of optimality, though not always formally so (Hockley and B. B. Murdock, 1987), and when formal, are usually predicated on RR maximization (Ditterich, 2006a; Deneve, 2012; Drugowitsch, Moreno-Bote, et al., 2012; Thura, Beauregard-Racine, et al., 2012; Moran, 2015). The main proposed reason for boundary collapse is between-trial variation in difficulty (Drugowitsch, Moreno-Bote, et al., 2012; Moran, 2015), though sometimes within-trial changes due to non-stationary accumulation (Ratcliff, 1980; Heath, 1982) or exogenously imposed time limits (Frazier and A. J. Yu, 2008; Karşılar et al., 2014) are also cited.

Fudenberg, Strack, and Strzalecki (2015) provide the first precise model of collapsing bounds derived as the solution to an optimization problem. Collapse in their model is driven by value-linked variation in drift rate that naturally accompanies economic decision making. As a result, the parameters that shape the threshold

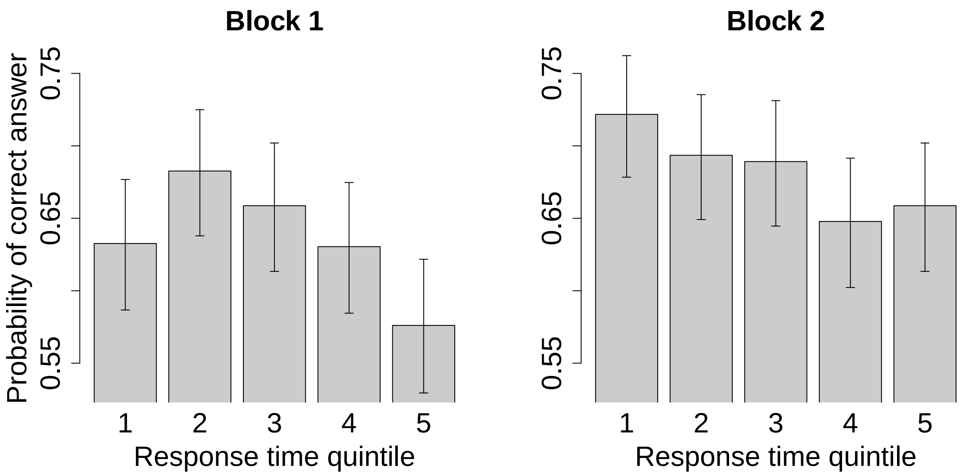

Figure 3.12: Relationship between accuracy and response time

are not arbitrarily free, but are functions of the primitive parameters defining the accumulation process.

For model comparison, I estimate a static threshold pure DDM with parameters representing decision threshold z and accumulation noise σ which are allowed to vary across blocks. The candidate parameters come from a 100x100 grid varying z from 50 to 446 by 4, and σ from 20 to 218 by 2, and the same simulation method is used.

Plotting choice accuracy grouped by response time quintile as in Figure 3.12 reveals a negative relationship between observed time and accuracy. This speed-accuracy complementarity is a basic feature of the data that indicates something is amiss according to the pure two-parameter DDM, which implies independence between observed time and accuracy. Complementarity is predicted by collapsing thresholds (Fudenberg, Strack, and Strzalecki, 2015) or by between-trial parameter variability (Laming, 1968; Ratcliff, 1978).

In order to formally compare models, I calculate the Bayesian Information Criterion (BIC) according to both models for every participant, $BIC = -2 \log L + m \log n$, where L is the maximum log-likelihood, m (= 4) is the number of free parameters in the model, and n (= 200) is the number of data points. The difference between BICs provides a measure of relative fit. In this case because the two models have the same number of parameters, the difference reduces to twice the difference in log likelihoods, but Bayesian interpretations of this value remain. In particular it is considered to approximate the Bayes factor, and Kass and Raftery (1995) provide recommendations for judging evidence strength accordingly. Figure 3.13 shows the difference in BICs for every individual in each block along with Kass and Raftery's (1995) benchmarks. Overall, there is compelling support for the uncertain-difference model, which is favored strongly or very strongly for 20 of 23 individuals in the full dataset. This is not universally true, however; considering the first block alone, the uncertain-difference model is favored strongly or very strongly for only 7 of 23 individuals, and the pure static model fits about the same for 6 individuals, if not better by a slim margin.

The BIC values can also be used to compute posterior model probabilities in another Bayesian analysis that accounts for uncertainty in model selection (Raftery, 1995; Wasserman, 2000). This is the central model comparison technique in the large-scale meta-analysis of Hawkins et al. (2015), and I employ it to weigh my results more directly against theirs. Supposing a uniform prior over the two competing models, the posterior probability of model j is

$$P(M_j|D) \approx \frac{\exp(-\frac{1}{2}BIC_j)}{\sum_k \exp(-\frac{1}{2}BIC_k)}.$$

I calculate these values for each individual based on data from the first and second blocks independently and combined. Figure 3.14 displays the results. Each stacked bar represents a single individual. The light and dark blue segments of each bar represent the posterior probabilities for the certain- and uncertain-difference models respectively. Participants are ordered by model probability separately in each subset of data. Overall, the data supports the uncertain-difference model extremely strongly, though less so in the first block alone, as before.

These analyses hint that the uncertain-difference model fits worse in the first block relative to the second block. This indeed seems to be the case, as shown in Figure 3.15, which plots the log-likelihoods for each individual in both blocks, and suggested by a paired t-test ($p = .065$). Such a difference could be due to greater experience —- which might familiarize participants with the task or facilitate sophistication in threshold setting —- or greater incentives —- which might boost their motivation to apply complex behavioral rules. Some additional evidence suggests the former. In response to a post-experiment question, five individuals (marked red in Figure 3.15) reported having taken part in a past experiment roughly similar to the present one. The uncertain-difference model fit this subset of participants better than others in the first block ($p = .027$, permutation test), but not in the second block

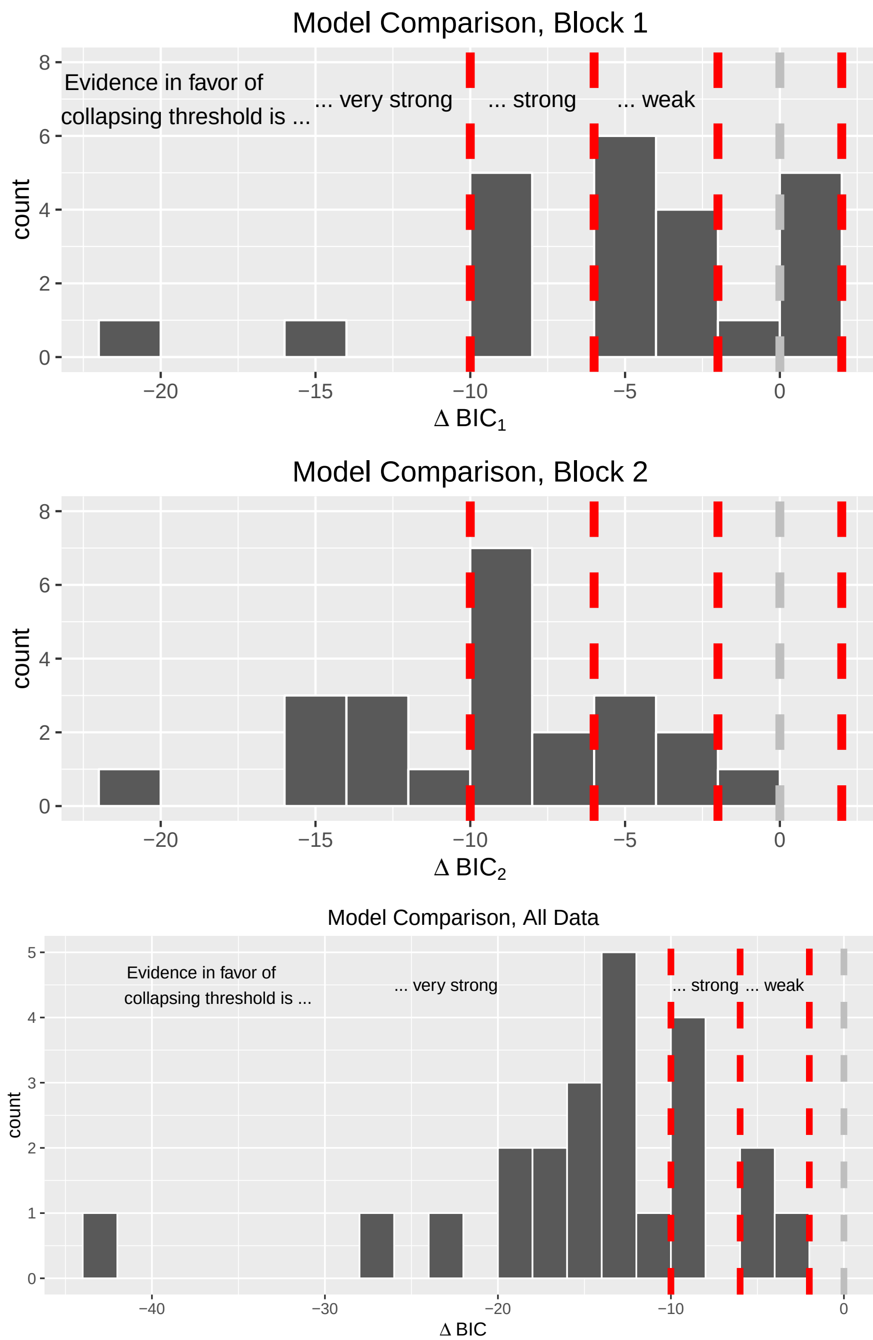

Figure 3.13: Comparisons of model fit for uncertain-difference collapsing bounds DDM versus pure DDM

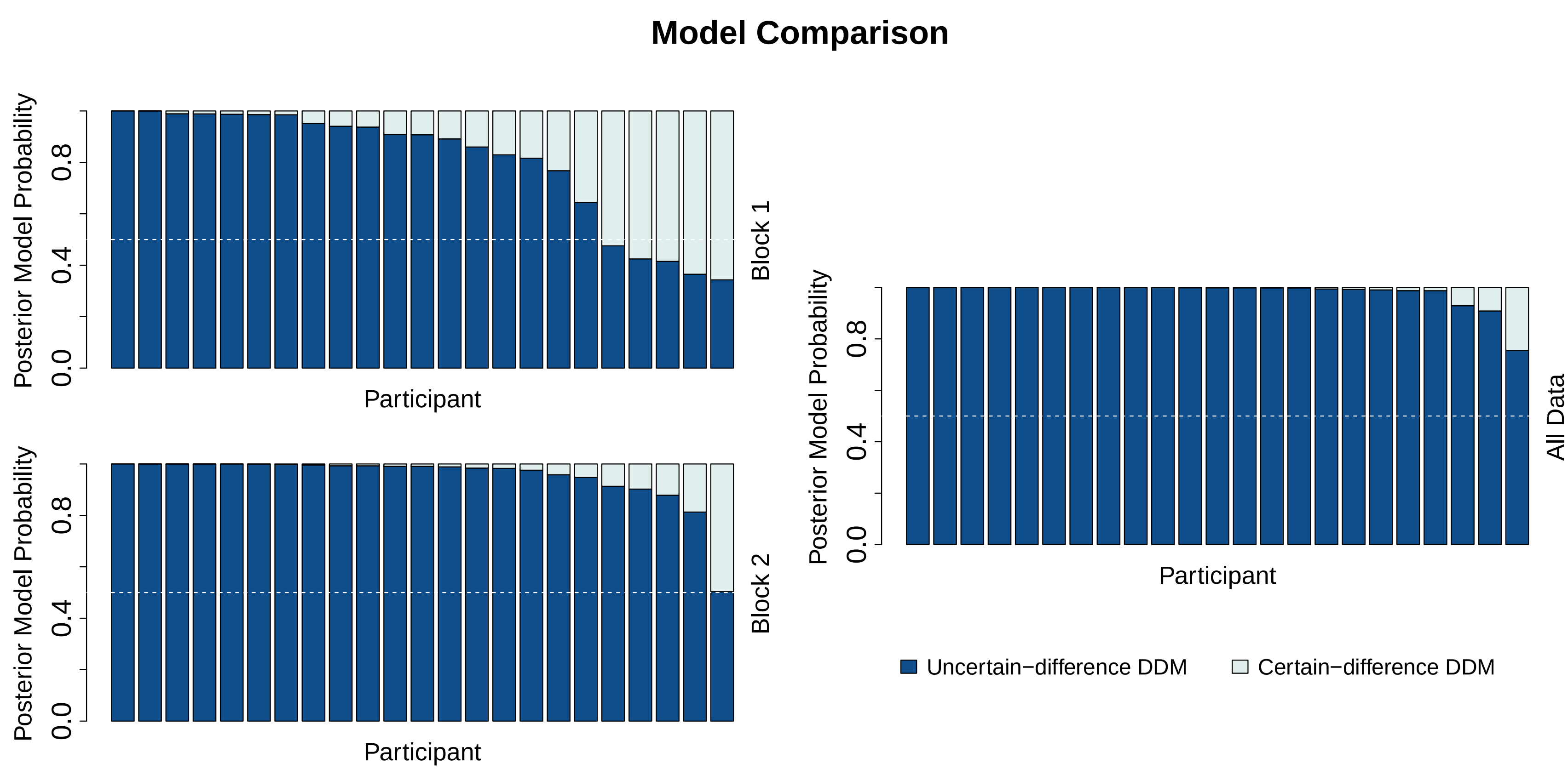

Figure 3.14: Bayesian comparisons of model fit for uncertain-difference collapsing bounds DDM versus pure DDM

(p = .421). This is consistent with basic task familiarity improving adherence to the predictions of a sophisticated model.

The uncertain-difference collapsing bounds model enjoys far more support in my data than the general collapsing bounds models do in the datasets assembled by Hawkins et al. (2015). There are a few possible reasons for this discrepancy.

The models I use are less parameterized than theirs, out of theoretical suitability and computational necessity. Their collapsing bounds DDM and simple DDM respectively contain up to 14 and 12 free parameters to allow full between-trial and between-condition parameter variability, while both of mine contain only 4. This may shift the scales in favor of their fixed bounds model for two reasons. First, their fixed bounds model contains additional parameters that are known to capture realistic features of observed data. For example, they allow for variance in the drift rate which could predict speed-accuracy complementarity. Thus the richness of their simple DDM may be critical in soaking up variation that might otherwise be attributed to boundary collapse. Second, their collapsing bounds model is penalized more heavily for its extra degrees of freedom. The uncertain-difference DDM runs on fewer parameters because its theoretical foundation connects threshold shape to the basic DDM parameters. The sharpness of the uncertain-difference DDM's predictions may be a major advantage that enhances model fit at minimal expense in terms of degrees of freedom.

In line with the second explanation and contrary to the first explanation, equalizing the number of free parameters in all of the Hawkins et al. (2015) models by eliminating between-trial variability for the more complex models increases the relative performance of the collapsing bound and urgency gating models. However, even with this equalization, the fixed bound model still performs at least as well in their analysis as the collapsing bound and urgency gating models combined. That is, their model with a static threshold and between-trial variability fits their data about the same as their models with collapsing bounds and no between-trial variability. Thus model degrees of freedom alone do not appear to fully explain the disparity in our findings.

Besides parametric considerations, my experimental paradigm may be better suited to elicit and measure threshold collapse. First, my random dot motion task involves a substantially finer grid of difficulty levels. Since left- and right-coherences are drawn from (rounded) Gaussian distributions with mean 25 and standard deviation 7, the difference between them – which constitutes the difficulty level, and is tied to the

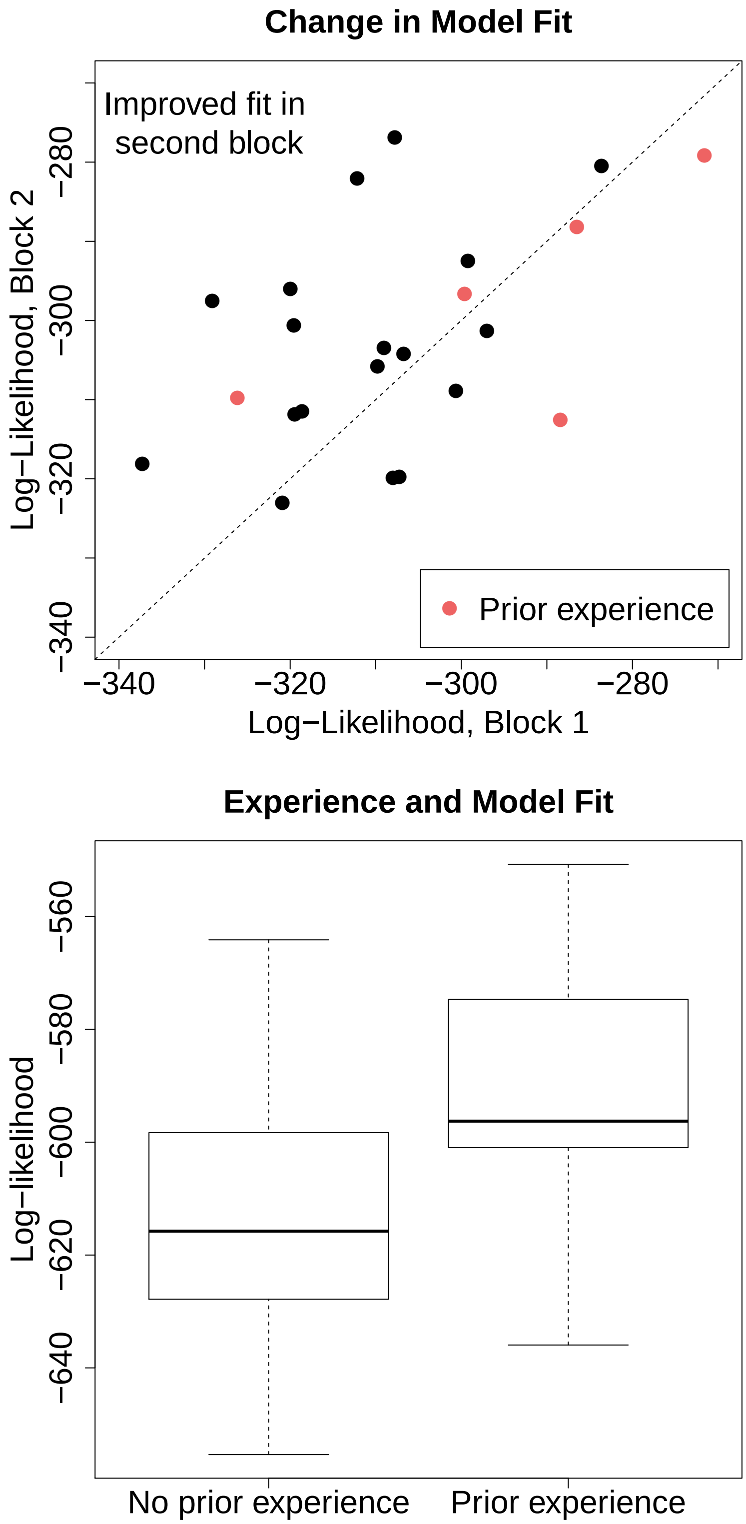

Figure 3.15: Experience and uncertain-difference DDM model fit

drift rate – is also Gaussian with mean 0 and standard deviation 14. This means the difference in coherences will take on magnitudes ranging from 0 to approximately 14×3 sd $= 42$ (in the data ranging from 0 to 38), which is six to ten times the number of levels used for each of the Hawkins et al. (2015) datasets. My model estimation does not collapse these levels into smaller, coarser bins, and instead fully exploits this extra order of magnitude of richness. In addition, half of the trials are expected to have coherence differences that lie between magnitudes 0 and 10. These trials generate relatively low drift rates that lead to high variation in response time, increasing the support of the data. Thus decisions both fast and slow are heavily sampled, as desired to identify collapse.

Decisions in my task also take longer than in most other perceptual experiments. For example, the 90th percentile response time was under 1.2 seconds in eight of the nine datasets analyzed by Hawkins et al. (2015) and under 1 second in five of the six experiments conducted by Voskuilen, Ratcliff, and P. L. Smith (2016). My experiment by contrast had a 90th percentile response time of nearly 8 seconds. This longer duration allows more scope for individual control over the decision rule. Finally, my participants are humans explicitly informed of the coherence distributions. Thus they can tap this prior information in the construction of their decision rule. Since these priors do not factor into simpler DDMs, any such information usage will imply sophistication in the decision mechanism.

3.4 Conclusion

Diffusion models of deliberative time allocation have become hugely popular over the last handful of decades. Part of their appeal lies in the optimal properties they inherit from efficient statistical rules (Abraham Wald and Wolfowitz, 1948; Arrow, Blackwell, and Girshick, 1949; Rafal Bogacz, E. Brown, et al., 2006). The extent to which observed deliberation can be grounded in optimization is an open question, but most attention to date has been focused on criteria that are not universally appropriate definitions of optimality. I lay out a method to test expected utility maximization in these settings that accommodates flexible characterizations of subjective individual preferences. This method combines versatile statistical tools to check for consistency of underlying preferences across different environmental conditions, in the spirit of economic tests of revealed preference. The prime virtue of the approach is its flexibility, while costs are paid in computational power and the need to specify how preferences should manifest in varying environments.

In conjunction, I run perceptual decision making experiments to empirically assess optimal deliberation by varying problem difficulty and incentive structure. I find that the hypothesis of optimality is rejected for about a third of the sample when difficulty changes, and for nearly everyone when monetary stakes change. Since consistency is a minimal condition for rationality, these figures should be taken as lower bounds on the sample prevalence of suboptimality conditional on the models studied. These results imply that existing theories of optimal deliberation are missing some meaningful piece of the picture. If the test rejects the null hypothesis, then either preferences are correctly specified and the agent is behaving inconsistently, or preferences are misspecified. More work is required to determine which is the case.

Further, one of my experiments is designed to investigate a new version of the DDM (Fudenberg, Strack, and Strzalecki, 2015) which constitutes the first precise model of collapsing decision thresholds developed as the solution to an optimization problem. I conduct the first empirical test of this theory and compare it to a standard static threshold model. The evidence strongly favors the new model, particularly among participants with task experience. This contrasts past studies in which collapsing thresholds fare worse than static thresholds. Strong theoretical foundations may be key in producing complex models that make precisely accurate predictions. Researchers are constructing versions of the drift diffusion model inspired by economic applications. Such settings tend to involve lengthier decisions and more scope for prior information to influence decision rules. These novel theories may generate fresh and valuable insights in predicting deliberative behavior.

Chapter 4

ECHOES OF THE PAST: ORDER EFFECTS IN CHOICE AND MEMORY

4.1 Introduction

Choices frequently must be made from options that appear in a sequence, such as when consumers decide between presented products, employers evaluate prospective job candidates, or judges appraise athletic or musical performances. A wide range of competitive scenarios are characterized by winner-take-all incentives. With much at stake, do outcomes reflect valid rankings of talent and value? In a consumer choice setting, I experimentally study how constraints on memory systematically bias preferences after sequential assessment, and how such biases may be alleviated based on principles of memory.

An empirical finding has been documented across several settings that a contender's chances of being selected are systematically related to their serial position, i.e. where they appear in the sequence of contenders. The very latest and the very earliest to appear are more likely to win compared to intermediate contenders. This is striking because if ordering is random, as is often explicitly the case in an attempt to be fair, then rank should be unrelated to serial position. These order effects have been found in the American Idol television franchise (L. Page and K. Page, 2010), the Eurovision Song Contest (Bruin, 2005; Bruin, 2006), the Queen Elisabeth music competition (Flôres Jr and Ginsburgh, 1996; Glejser and Heyndels, 2001), and across figure skating (Bruin, 2005; Bruin, 2006), gymnastics (Rotthoff, 2015), synchronized swimming (Wilson, 1977), sales presentations (Wagner and Klein, 2007), environmental policy evaluation (Payne et al., 2000), and political ballots (Meredith and Salant, 2013), with significant consequences for competitors' careers (Ginsburgh and Van Ours, 2003). Moreover, these effects seem to apply to choice among consumer products where option attributes are objectively fixed and order is experimentally randomized (Li and Epley, 2009; Mantonakis et al., 2009; Schosser, Trarbach, and Vogt, 2013).

Order effects are often hypothesized to occur in large part due to properties of memory. In the oldest memory paradigm used in psychology (beginning with Ebbinghaus, 1885), lists of items (usually words) are presented to people who are later asked to recall them. The earliest and the latest items in these lists tend to be best remembered. Serial position effects have been studied extensively and found across a variety of materials and timescales. Such primacy and recency effects in memory clearly parallel those in judgment. If memory limitations are responsible for serial position effects in judgment, then principles of the former discovered over the last century should predict when the latter will obtain and how strong they will be. These principles can accordingly inform us about how to shape or de-bias decision making.

Primacy and recency effects are believed to occur for different reasons. The earliest items are favored in both memory and choice because judges can mentally rehearse them more than later items, while the latest items are favored because they remain fresh in judges' memories. Intermediate items are weaker in memory because they do not benefit from either of these forces and thus, as the hypothesis goes, their mental impact fades and they receive worse evaluations. If these hypotheses are true, the implications can be very counterintuitive. In particular, they entail that interviewers who spend all available time deliberating and customers who are especially attentive and involved will be *most* susceptible to these biases. Bias can therefore be *reduced* by *impairing* an agent's mental activity.

Various kinds of evidence in the memory literature support these claims. Some experimental research explicitly tracks participant rehearsal patterns by asking participants to say out loud what comes into their minds during memorization, known as an "overt rehearsal" paradigm. These studies find that early items are indeed rehearsed disproportionately often (e.g. Rundus, 1971; Brodie and B. B. Murdock, 1977; B. Murdock and Metcalfe, 1978; Tan and Ward, 2000; Ward, 2002). Many other studies use what is called a "continuous distractor" paradigm in which a distracting task occurs in between item presentations. This body of work has found that when the distractor task prevents rehearsal, primacy effects deteriorate while recency effects remain (e.g. Bjork and Whitten, 1974; Glenberg, Bradley, Stevenson, et al., 1980; Watkins, Neath, and Sechler, 1989; Howard and Michael J Kahana, 1999; Michael Jacob Kahana, 2012). Such results obtain across a wide range of inter-stimulus intervals, from tenths of seconds (Neath, 1993) to days (Glenberg, Bradley, Kraus, et al., 1983).

I conduct the first experiment testing the prediction that cognitive load inhibits choice-based primacy effects by altering opportunities for memory consolidation. I study the effects of two kinds of load which should reduce the ability or willingness of participants to rehearse items, disproportionately hindering those presented early on. One source of load is a distraction in between stimulus presentations, and the other is fatigue as the experiment progresses. In the experiment, participants are shown sequences of digital art, after which they report their favorite, and take a memory test. Thus in contrast to other experiments, I measure memory and choice together to directly connect the two. I also observe the impact of artificially- and naturally-occurring cognitive load. For half of the experimental blocks, a distracting task occurs before and after each art piece is presented, in line with the continuous distractor paradigm.

I observe strong primacy and recency effects for both choice and memory in non-distractor trials early in the experiment, when participants would be most willing and able to engage in rehearsal. However, the primacy effect is diminished in both domains by the distractor task, and in the non-distractor trials in the later part of the experiment. These findings indicate that disrupting encoding by increasing cognitive load is a useful principle upon which to build interventions. When people are less able or less inclined to rehearse items, they will exhibit smaller decision primacy effects without disturbing recency effects.

Despite the appeal of the overall idea, not much work has focused on tying memory to judgment in this context. Kardes and Herr (1990) found that when participants were motivated to remember the sequentially-presented attributes of two options, the early presented attributes were weighed more heavily in choice and were better recalled. Mantonakis et al. (2009) observed that the last wines tasted in a sequence were preferred more only in longer sequences, when memory would be under a greater load. Li and Epley (2009) showed that increasing the delay between presentation and evaluation of a desirable individual painting led to lower evaluations. However, even the existence of primacy versus recency effects in judgment varies substantially across studies. Such inconsistencies could be reconciled based on subtle situational differences thought to influence memory strength. For example, despite having apparently similar setups, Mantonakis et al. (2009) reported both primacy and recency effects, while Li and Epley (2009) observed only recency effects. This may have been caused by procedural differences, as participants in the former study merely waited for a period of time in between stimulus presentations while participants in the latter study engaged in filler tasks.

The current project also contributes to a broader research program in neuroeconomics. Many studies in neuroeconomics concentrate on choices stemming from habitual learning, with experiments that involve repeated presentation of a few stimuli arbitrarily linked to value. However, evidence is emerging of decision making based on alternative systems that invoke memory for items that are only presented once (e.g. Wimmer and Shohamy, 2012; Wimmer, Braun, et al., 2014). Such systems may be more applicable when choices must be made with limited direct experience and associations must be flexibly made over extended periods of time. My investigation centers on a phenomenon stemming from memory for trial-unique stimuli. I produce data on preferences in a paradigm that taxes memory over longer intervals than standard reinforcement learning studies. My work may thus be a useful behavioral stepping stone for future neuroeconomic analyses.

4.2 Experimental Design

This experiment was designed to study the interaction between memory and judgment. To connect existing bodies of work as directly as possible, I retained the central design elements of related memory and judgment paradigms. Participants were shown sequences of art pieces and subsequently selected their favorite, similar to Li and Epley (2009). However, in some blocks they were given a distractor task between each stimulus presentation, according to the continuous distractor paradigm (Bjork and Whitten, 1974). After each sequence, they were also given a recognition memory test to assess the strength of their memories. I assess the consequences of two kinds of cognitive load on the primacy effect: one imposed by the distractor task, and the other resulting from fatigue. Both should reduce the ability or willingness of participants to engage in rehearsal that would disproportionately benefit early options.

The setup is depicted in Figure 4.1. In each of 36 blocks, participants were shown a sequence of 5 abstract digital art images, each displayed for 5 seconds and shown in only a single block. The images were chosen to be complex visual stimuli which must be evaluated holistically. Thus, importantly for present purposes, the exact nature and evoked feeling of each image cannot be perfectly remembered. Responding to a post-experiment question about strategies used to remember the images (see Appendix C), participants did indeed report condensing images to their attributes such as color and shape or associating them with evocative names.

After each sequence, participants were asked to indicate which image they liked the most. They did so by hitting the number key corresponding to its serial position, i.e. hitting "1" for the 1st image shown, "2" for 2nd image, and so on. They could

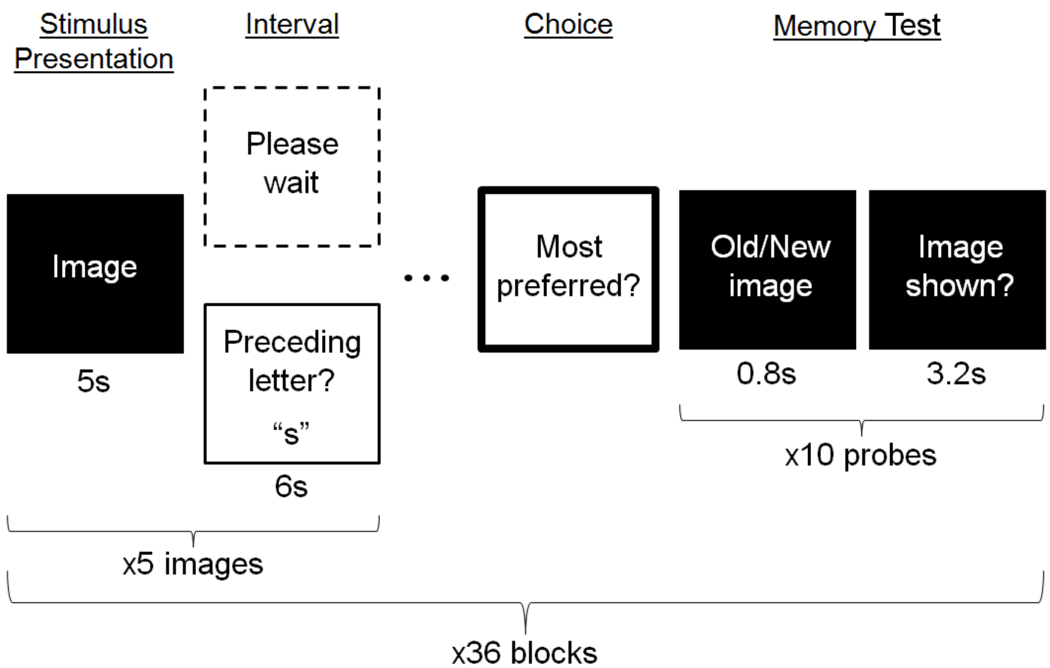

Figure 4.1: Experimental setup

also hit "0" to represent indifference between all 5. This indifference option serves two purposes. First, it provided a convenient default to ensure choices of 1 through 5 were intentional. Fortunately, only a small minority (7%) of choices were of this option after removing a single individual who selected it exclusively. Second, it enables me to measure a possible cost of cognitive load. Cognitive load might distract participants, impairing their ability to evaluate and discriminate between options. This impairment could otherwise lead to greater noise in people's choices, masking the hypothesized effects.

Before and after every image in a block, participants faced either timed delays or distractor task trials. Half of the blocks (randomly chosen) exclusively involved delays and the other half exclusively involved distractors. The timed delay consisted of an otherwise black screen saying "Please wait" for 2 seconds followed by a pure black screen for 4 seconds. The distractor task was styled after existing studies of order effects in psychology, which normally include simple arithmetic problems such as 23 + 55 (e.g. Bjork and Whitten, 1974). To avoid possible anchoring effects that might bias valuation, I instead used a task wherein participants were shown a random letter on the screen and responded by hitting the key of the alphabetically preceding letter (e.g. if "r" was shown, they should hit "q"). Three of these trials occurred in every interval, separated by a 0.3 second fixation cross. They had to respond within 3 seconds and earned \$0.01 for a correct answer. Participants on average achieved over 80% accuracy. Distractor trials were closely matched with the delays in terms of elapsed time. On average, distractor intervals took 5.61 seconds, with 80% taking between 4.12 and 7.28 seconds, close to the 6 seconds of empty delays.

After the image selection occurred, participants engaged in a recognition memory test to gauge memory strength. In every block, they were shown 10 blurred images in random order, each on screen for 0.8 seconds after a 0.5 second fixation cross. Five of these were versions of the images just displayed in the preceding sequence, while the rest were not shown at all in the experiment. After each image, participants had to indicate whether or not they had seen it before, and if they had, in which position it was displayed. They hit the key "0" if they believed the image was new, or a key from "1" to "5" to note its position if they believed they had encountered it previously. Although serial memory paradigms tend to test memory by free recall, this was not possible given the kind of stimulus used. Participants had to respond within 4 seconds and earned \$0.04 for a correct answer. In line with previous memory research, many participants explicitly stated that they rehearsed the items in order of appearance. Several also noted that the distractor task did indeed impair their ability to do so, and some informally admitted that they became tired or bored in the later part of the experiment.

This group of tasks was repeated 36 times for each individual (on top of a sample block to familiarize participants), and half of these blocks (randomly dispersed) involved distractors. The experiment was long (60-–75 mins) and fatigue-inducing due to the large number of decisions that needed to be made in short periods of time. Participants were 69 individuals recruited through the Caltech SSEL. In the following analyses, two participants are excluded, one who selected the indifference option in every block and the other who wrote notes on paper as a memory aid.

4.3 Results

As expected, substantial order effects are observed in the behavioral data. The aggregate distribution of stated preferences in the first half of the no-distractor blocks —- i.e. under no cognitive load — is shown in Figure 4.2 with 95% confidence intervals for proportions. Both primacy and recency effects are strongly present, and a chi-square test rejects uniformity of the distribution ($p = 1 \times 10^{-5}$, omitting the indifference option). This result replicates previous findings of order effects in

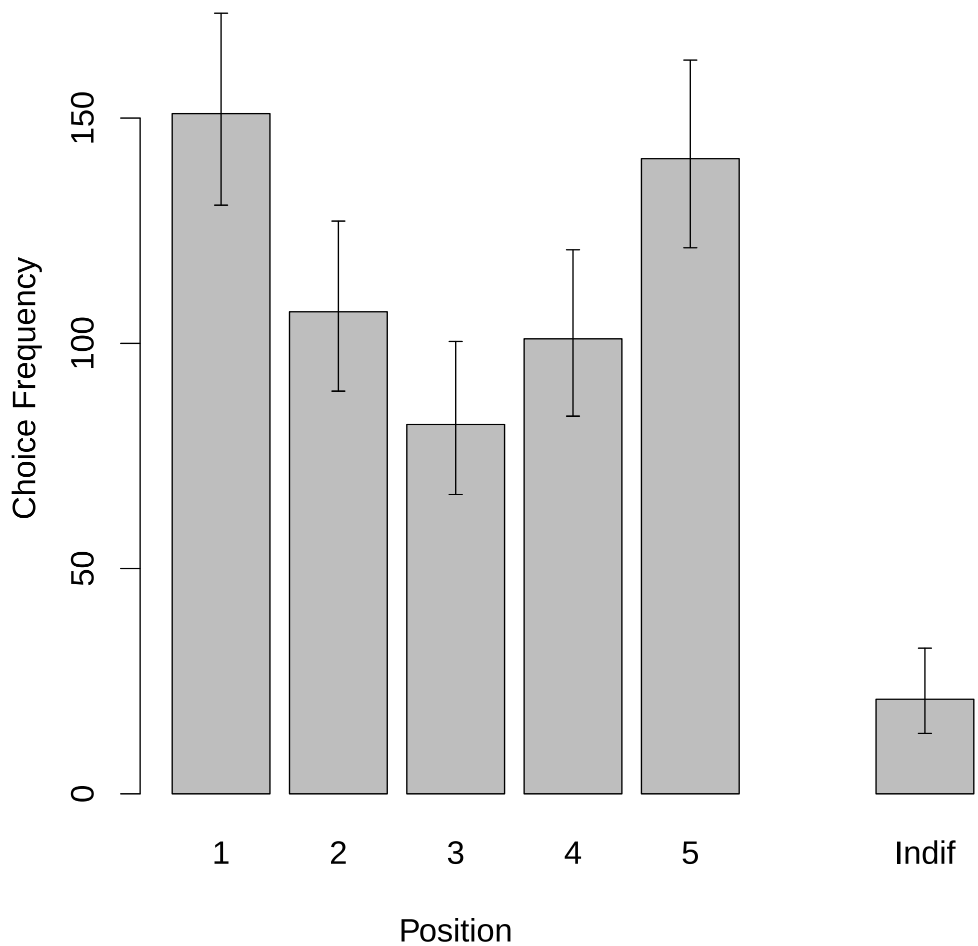

Figure 4.2: Baseline order effect in choice

judgment (Mantonakis et al., 2009; Li and Epley, 2009). Notice that these previous experiments present each participant with only a single sequence of items. Thus primacy effects in their studies could in principle be caused by a novelty bias, meaning that the earliest item is favored because it is a novel stimulus type to which people quickly become accustomed. Such a confound does not apply here.

Figure 4.3 depicts the effects of cognitive load from the distractor task and fatigue. It compares the choice distribution under no load to the distractor blocks in the same, first, half (holding fatigue fixed) and the second-half no-distractor blocks (holding distraction fixed). These distributions appear to be significantly different from the no-load baseline. In particular, there seem to be selective reductions in the primacy effect; the advantage of the first option relative to intermediate options erodes.

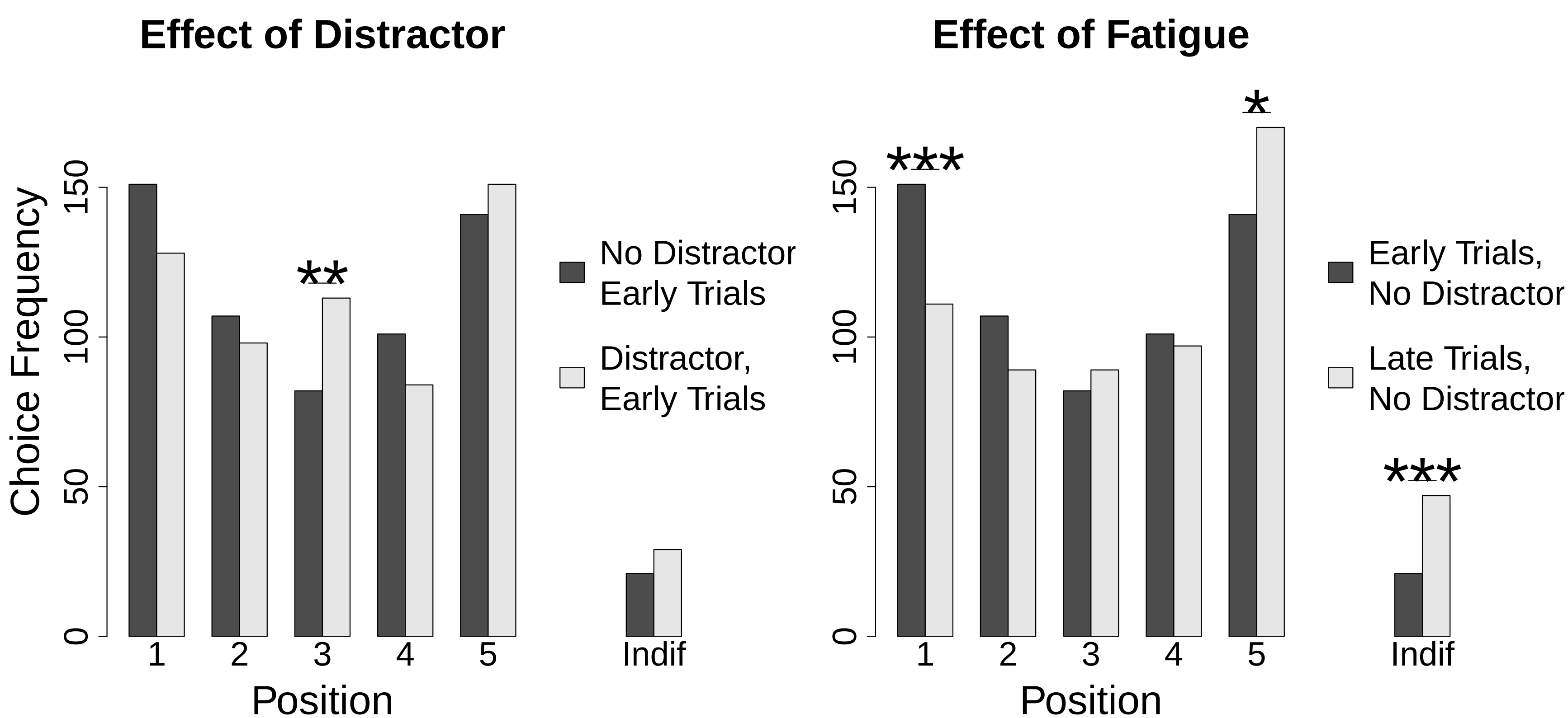

Figure 4.3: Changes in order effect due to cognitive load

I analyze the primacy effect in depth. To make the trends more apparent, I also display the data with blocks grouped into sets of 3. Figure 4.4 depicts the aggregate probability of choosing the first option, Figure 4.5 depicts the overall accuracy and average response time on the memory test for the first option, and Figure 4.6 depicts the overall accuracy and average response time on the distractor task.

In every block in the first half of the experiment, the probability of choosing the first option under no load is higher than or at least as high as under distraction. The overall choice frequency for the first position is 25% at baseline (significantly different from the uniform probability of 20%; $p = .002$, Z-test) compared to 21% in distractor blocks (not significantly different from 20%; $p = .482$, Z-test). This reduction in the primacy effect is confirmed by a paired t-test comparing the choice probabilities within blocks ($p = .009$). The same result obtains with a more robust nonparametric permutation test based on the same t statistic, which randomly relabels the data under the null hypothesis that the distractor task has no effect on choice frequency ($p = .009$). This analysis effectively controls for within-block variation. Intriguingly, controlling also for individual identity markedly weakens the effect (yielding $p = .097$, permutation test), suggesting high levels of individual heterogeneity, and indicating that the distractor effect is not very consistent on a within-individual basis. People also respond to first items more slowly and less accurately on the memory test in distraction blocks ($p = .004$, t-test; $p = .004$, permutation test), and exhibit below-average accuracy on the distractor task itself ($p \ll .001$, Z-test for equality of proportions across experimental halves).

In the second half of the experiment, the probability of choosing the first option under no load drops significantly compared to the first half, from 25% to 18% (not significantly different from 20%; $p = .354$, Z-test). This is confirmed by a Z-test for equality of proportions ($p = .006$), and the same result obtains with a permutation test based on the same Z statistic, which randomly relabels the data under the null hypothesis that fatigue has no effect on choice frequency ($p = .004$). In this case, controlling for individual identity makes virtually no difference in the effect's statistical significance (yielding $p = .005$, permutation test). Consistent with a drop in motivation near the experimental midpoint, accuracy in the memory test falls ($p = .028$, Z-test; $p = .020$, permutation test) and accuracy in the distractor task stops improving and appears to start gradually declining.

Figure 4.3 also reveals a potential downside of cognitive load. Although load reduces primacy effects, in the process of hindering cognition it may reduce people's ability to evaluate options. This side effect can be measured by the frequency with which people choose the indifference option. While fatigue seems to be associated with a significant rise in indifference (3.5% vs 7.8%; p = .002, Z-test for equality of proportions), the distractor task does not seem to produce any such change (3.5% vs 4.8%; p = .312). Hence, at least as perceived by individuals themselves, the distractor task does not appear to reduce their ability to discriminate between options.

In general, if order effects in choice are driven by differences in memory, comparable effects should be seen in both domains. Memory strength under no cognitive load is displayed in Figure 4.7 for each position, paralleling Figure 4.2's depiction of choices. Similar to choice behavior, memory test accuracy and response time exhibit a "U" (or inverted-"U") shape in the aggregate. In other words, participants answer both quickly and accurately when prompted with early and late items, and slowly and inaccurately with intermediate items. To ensure that the variation in memory is not due to chosen items being better remembered, I also plot memory strength split by position for the subset of items which were not chosen in each block. This depicts the relationship between position and memory undistorted by choice. The same "U" shape remains, indicating a direct causal pathway from position to memory.

The recency effect appears to be attenuated in both cases, though. This is in line with past memory research, and in the present work relates to a natural limit on experimental design. Recency effects are known to be reduced when the memory test is delayed beyond the end of the item study period (Bjork and Whitten, 1974; Poltrock and MacLeod, 1977; Glenberg and Kraus, 1981; Talmi and Goshen-Gottstein, 2006). In this experiment, the assessment of decision making was prioritized over the measurement of memory. Therefore the choice period occurred immediately after item presentation, forcing the recognition test to be delayed as a consequence. Thus, as a caveat, the recognition test may not be ideal for capturing recency effects, but should adequately portray primacy effects.

Figure 4.8 depicts how cognitive load modulates the impact of position on recognition test accuracy and response time. (The last position is grayed out to signify that it is an inaccurate measurement of the underlying phenomenon.) As above, this comprises the subset of data for which the position was not chosen. The accompanying regressions (which exclude data for the last position) reported in Table 4.1 predict memory test performance by the position of the item, and whether the block was a distractor block and in the second half of the experiment. The dis-

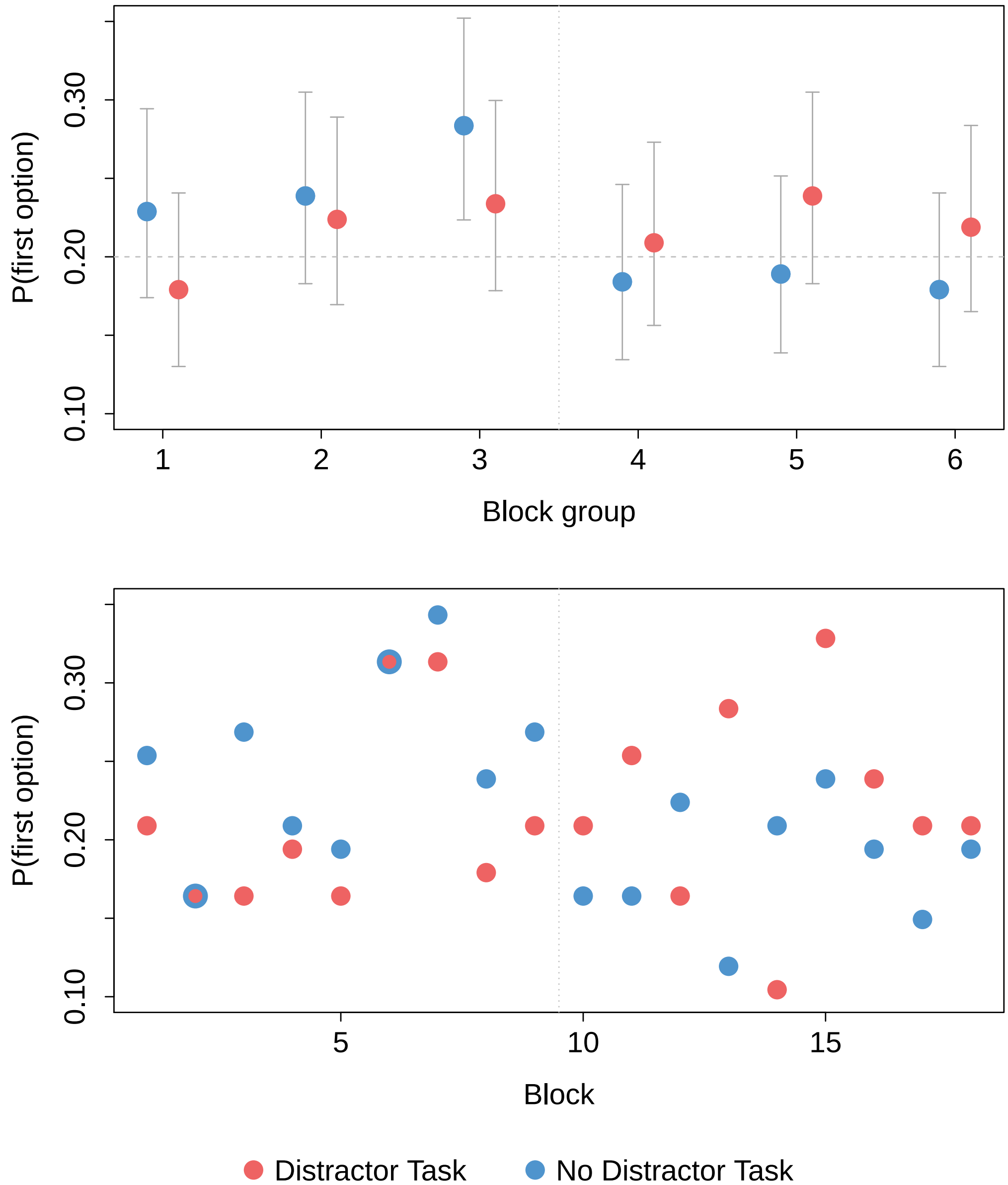

Figure 4.4: Changes in primacy effect due to cognitive load. Top row: blocks grouped into sets of three. Bottom row: individual blocks.

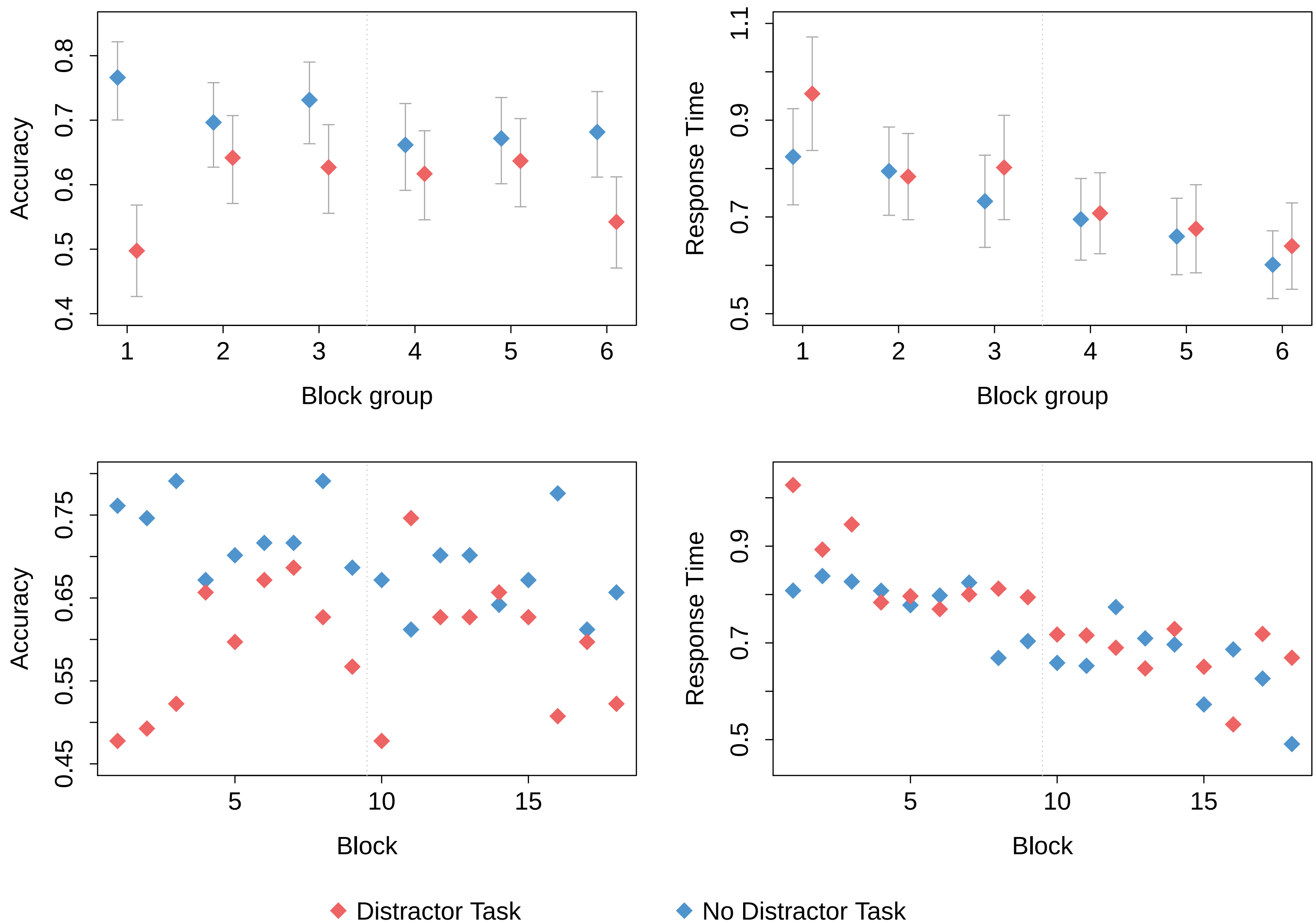

Figure 4.5: Changes in memory for first option due to cognitive load. Top row: blocks grouped into sets of three. Bottom row: individual blocks.

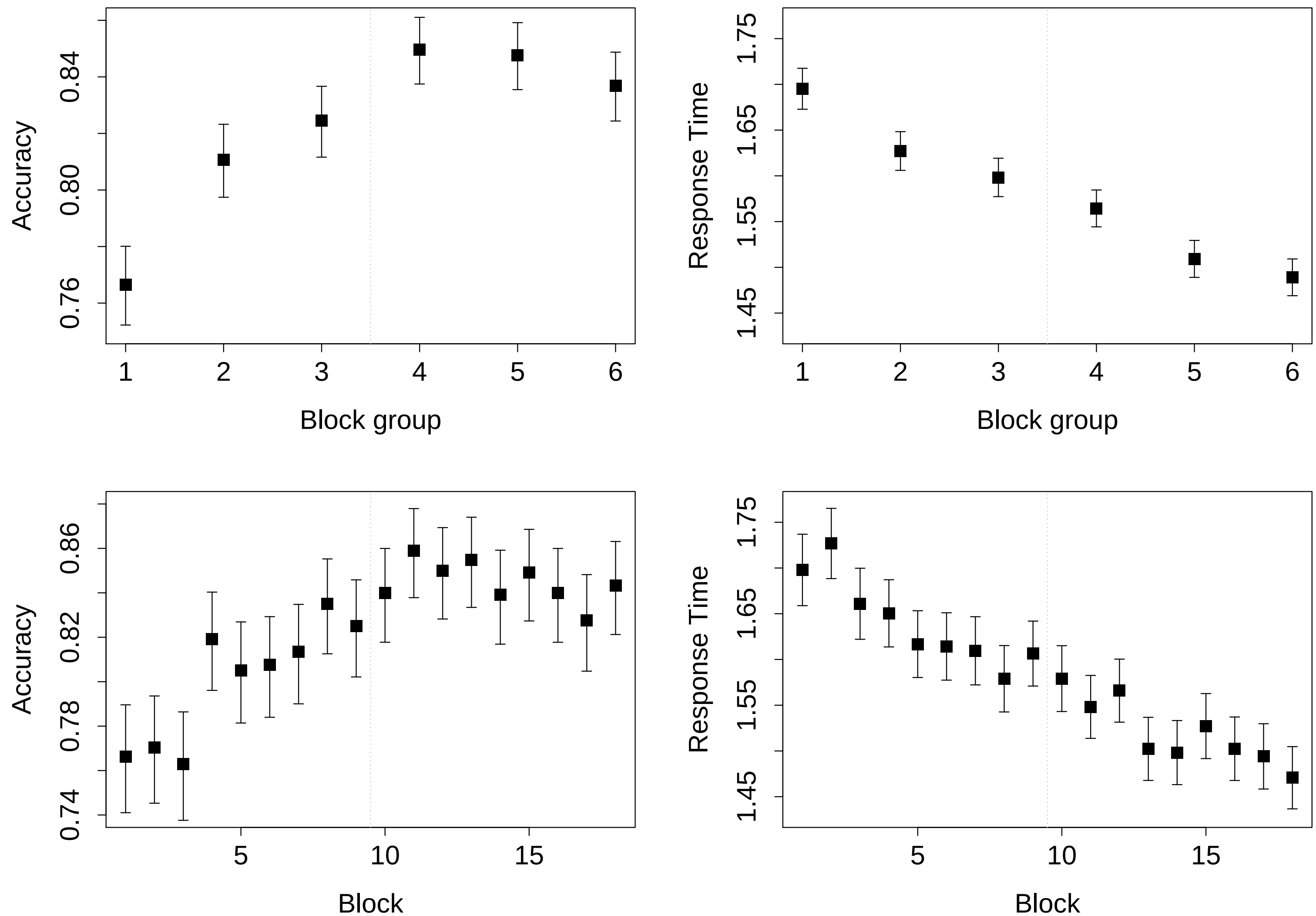

Figure 4.6: Changes in distractor task performance. Top row: blocks grouped into sets of three. Bottom row: individual blocks.

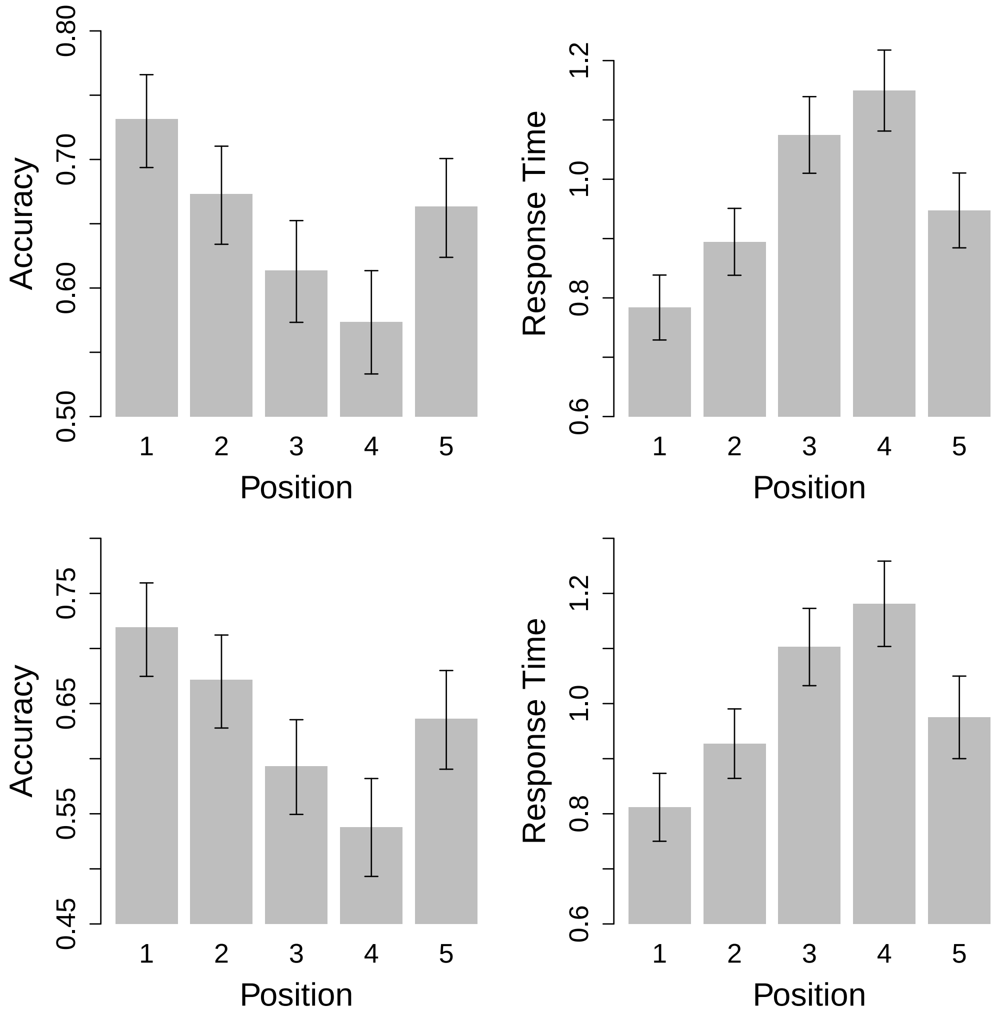

Figure 4.7: Baseline order effect in memory. Top row: all data. Bottom row: excluding data from items chosen in same block.

tractor task disproportionately slowed down responses to images that were shown in earlier positions, indicated by the negative coefficient on the interaction between distractor and position, though it has no comparable statistically detectable effect on accuracy. Similarly, fatigue disproportionately reduces accuracy on images that were shown in earlier positions, indicated by the positive coefficient on the interaction between block location and position, though it has no comparable statistically detectable effect on response time. These trends can also be observed in Figure 4.8. Cognitive load overall thus seems to modulate the effect of position on memory, disproportionately impairing the earliest items as expected.

Having established a link between position and memory, I look at the link between

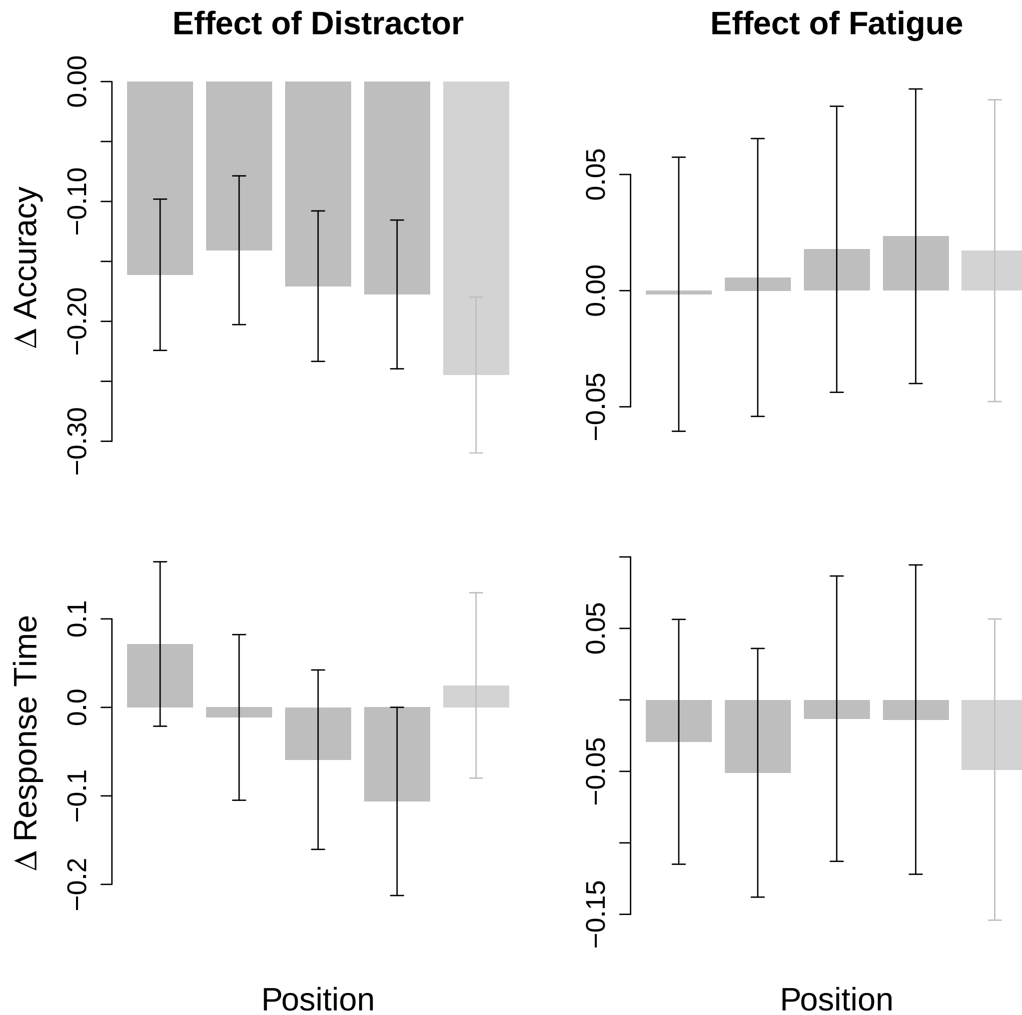

Figure 4.8: Effect of cognitive load on memory × position interaction

memory and choice. Although memory strength consists of absolute measurements, choice is relative, so I construct an individual-level index of relative memory strength with which to predict choice probabilities. I define this index for each position to be the test accuracy of items presented in that position normalized by the sum of accuracies for items in all positions. Thus, for example, if a person responded to the first position correctly 8 times out of 10 and responded to all positions correctly 30 times out of 50 in total, their relative memory strength for the first position would be $8/30 = 0.267$, while if they responded correctly to 20/50 in total, it would be $8/20 = 0.400$. This calculation transforms the independence of memory into the comparative nature of choice.

I regress choice probabilities on relative memory strength for each position and

Table 4.1: Position-linked effect of cognitive load on memory strength

	Dependent variable:	
	Memory Test Performance	
	Accuracy	Response Time
Constant	1.725***	1.103***
	(0.107)	(0.027)
Distractor	−0.675***	0.112**
	(0.193)	(0.052)
Second Half	−0.043***	−0.010**
	(0.013)	(0.004)
Distractor × Second Half	0.028	−0.003
	(0.017)	(0.005)
Distractor × Position	−0.036	−0.054**
	(0.062)	(0.022)
Second Half × Position	0.012***	−0.001
	(0.004)	(0.002)
Distractor × Second Half × Position	−0.005	0.0003
	(0.006)	(0.002)
Individual and Position Fixed Effects	Yes	Yes
Note: SEs clustered by individual.	*p<0.1; **p<0.05; ***p<0.01	

individual in each condition. The data is plotted in Figure 4.9. If there were no systematic relationship between memory and choice, then the only correlation between the variables would be due to noise. The regression in Table 4.2 reveals a strong positive relationship, although the standard errors are inappropriate due to adding-up constraints; a positive relationship between memory strength and choice for one given position will imply a positive relationship for other positions, artificially amplifying the correlation. A permutation test in which the relative memory strengths are randomly reshuffled for each individual corrects for this by constructing the test statistic distribution without the same parametric assumptions, and indicates the observed magnitude would be extremely unlikely under the null hypothesis that memory and choice are not connected (p = .0001). In conjunction

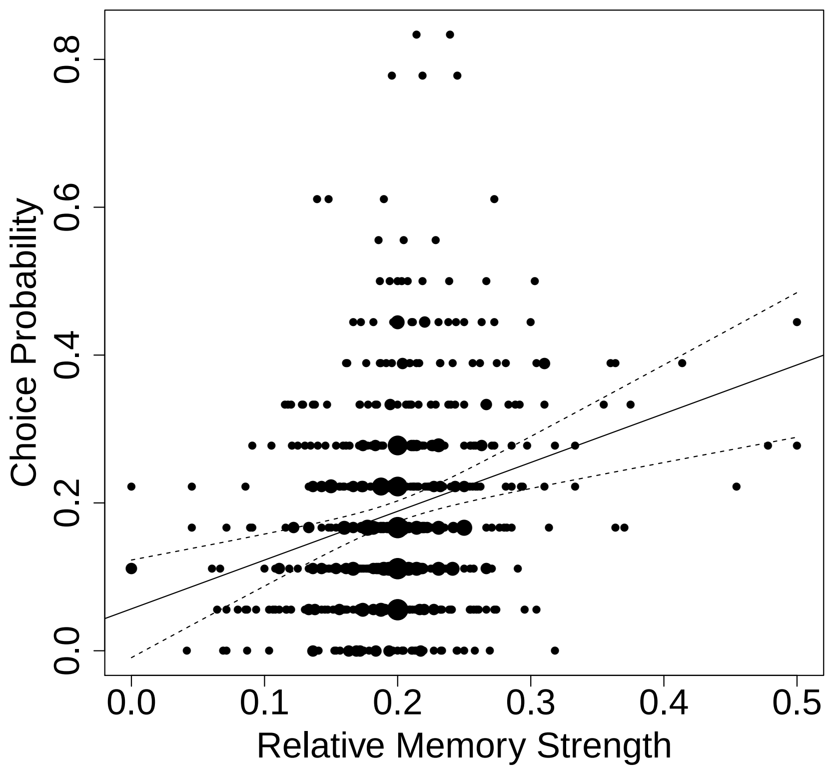

Figure 4.9: Relationship between memory strength and choice probability

Table 4.2: Estimated relationship between memory strength and choice probability

	Dependent variable:
	Choice Probability
Constant	0.057**
	(0.026)
Relative Memory Strength	0.660***
	(0.132)
Distractor	0.019
	(0.032)
Relative Memory Strength × Distractor	−0.109
	(0.158)
Note: SEs clustered by individual.	*p<0.1; **p<0.05; ***p<0.01

with the link between position and memory, this positive correlation provides some evidence that position causally influences choice through a path modulated by memory.

4.4 Conclusion

When people choose from options that are presented in a sequence, they tend to prefer the earliest and latest items. Such order effects have been found across a range of decision contexts, and resemble serial position effects found in studies of memory. If memory and choice are indeed connected, then factors influencing the former should also affect the latter. I conduct an experiment that connects design elements of choice and memory tasks in order to speak as directly as possible to the connection between domains. People indicated preferences over pieces of art presented to them sequentially. In addition to observing large order effects, I find that cognitive load, imposed by a distractor task and by natural fatigue, appears to substantially reduce primacy effects. This change is targeted and does not appear to weaken recency effects. Thus contextual factors or interventions that impair the ability or willingness of decision makers to rehearse options can selectively alter order effects.

This research has implications for contest organizers or employers trying to amend their selection procedures to select the best candidate, as well as for marketers curating product presentation in order to guide consumer choice. Minor procedural differences could remain unnoticed in many scenarios but nevertheless substantially alter overall outcomes. Consider for instance that there are commonly breaks for deliberation between sequential interviews. Despite enabling more information processing, such breaks may actually make decisions more biased due to an imbalance in which information can be processed. The kind of bounded rationality described here is natural and difficult to alleviate, but, counterintuitively, increasing the cognitive burden on agents can lead to more accurate evaluation and better decisions. This knowledge is useful because cognitive processing is in general far more easily impaired than improved. It opens up a suite of small, practical interventions that may prove useful in many important settings. Since order effects in memory have been found over a vast range of timescales, these insights could be widely applicable.

More broadly, bounded memory could also influence sequential search, in which decision makers themselves control how many options or how much information they are exposed to (e.g. Salant, 2011; Mogilner, Shiv, and Iyengar, 2013). Especially when people are under cognitive load, good alternatives in the past may be penalized because they are not recalled with full force. These options would be undervalued. People may also stop searching earlier than they otherwise would, since the memory-distorted historical value distribution would be depressed. Theoretical work building on these themes could generate novel predictions about how features of memory alter sequential judgment.

Decision theoretic models generally assume choices are made from elements in a mathematical set, in which item order is by definition irrelevant. This simplification is theoretically useful, but it is clear that order often matters in practice. The idea that value can be generated from various kinds of memory that may decay or be transformed with experience is straightforward. However, much remains to be discovered about the mechanisms that give rise to such evaluations. Memory is among the earliest pillars of experimental psychology, and exhibits many unusual and counterintuitive properties. Research in decision making has much to learn from over a century of accumulated knowledge regarding memory.

BIBLIOGRAPHY

Abdellaoui, Mohammed and Emmanuel Kemel (2014). “Eliciting Prospect Theory When Consequences Are Measured in Time Units:“Time Is Not Money””. *Management Science* 60.7, pp. 1844–1859.

Abeler, Johannes et al. (2011). “Reference points and effort provision”. *The American Economic Review* 101.2, pp. 470–492.

Allen, Eric J et al. (forthcoming). “Reference-dependent preferences: Evidence from marathon runners”. *Management Science*.

Anatolyev, Stanislav and Grigory Kosenok (2005). “An alternative to maximum likelihood based on spacings”. *Econometric Theory* 21.02, pp. 472–476.

Arrow, Kenneth J, David Blackwell, and Meyer A Girshick (1949). “Bayes and minimax solutions of sequential decision problems”. *Econometrica*, pp. 213–244.

Balci, Fuat et al. (2011). “Acquisition of decision making criteria: reward rate ultimately beats accuracy”. *Attention, Perception, & Psychophysics* 73.2, pp. 640–657.

Bartling, Björn, Leif Brandes, and Daniel Schunk (2015). “Expectations as reference points: Field evidence from professional soccer”. *Management Science*.

Bell, David E (1985). “Disappointment in decision making under uncertainty”. *Operations research* 33.1, pp. 1–27.

Bitzer, Sebastian et al. (2014). “Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model”. *Frontiers in human neuroscience* 8.

Bjork, Robert A and William B Whitten (1974). “Recency-sensitive retrieval processes in long-term free recall”. *Cognitive Psychology* 6.2, pp. 173–189.

Bode, Stefan et al. (2012). “Predicting perceptual decision biases from early brain activity”. *The Journal of Neuroscience* 32.36, pp. 12488–12498.

Bogacz, Rafal, Eric Brown, et al. (2006). “The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks.” *Psychological review* 113.4, p. 700.

Bogacz, Rafal, Peter T Hu, et al. (2010). “Do humans produce the speed–accuracy trade-off that maximizes reward rate?” *The Quarterly Journal of Experimental Psychology* 63.5, pp. 863–891.

Bowman, Nicholas E, Konrad P Kording, and Jay A Gottfried (2012). “Temporal integration of olfactory perceptual evidence in human orbitofrontal cortex”. *Neuron* 75.5, pp. 916–927.

Britten, Kenneth H et al. (1992). “The analysis of visual motion: a comparison of neuronal and psychophysical performance”. *The Journal of Neuroscience* 12.12, pp. 4745–4765.

Brodie, Delbert A and Bennet B Murdock (1977). “Effect of presentation time on nominal and functional serial-position curves of free recall”. *Journal of Verbal Learning and Verbal Behavior* 16.2, pp. 185–200.

Brown, Joshua W et al. (2008). “Relation of frontal eye field activity to saccade initiation during a countermanding task”. *Experimental Brain Research* 190.2, pp. 135–151.

Brown, Scott and Andrew Heathcote (2003). “QMLE: Fast, robust, and efficient estimation of distribution functions based on quantiles”. *Behavior Research Methods, Instruments, & Computers* 35.4, pp. 485–492.

Bruin, Wändi Bruine de (2005). “Save the last dance for me: Unwanted serial position effects in jury evaluations”. *Acta psychologica* 118.3, pp. 245–260.

– (2006). “Save the last dance II: Unwanted serial position effects in figure skating judgments”. *Acta psychologica* 123.3, pp. 299–311.

Busemeyer, Jerome R and James T Townsend (1993). “Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment.” *Psychological review* 100.3, p. 432.

Camerer, Colin et al. (1997). “Labor supply of New York City cabdrivers: One day at a time”. *The Quarterly Journal of Economics* 112.2, pp. 407–441.

Cheng, RCH and NAK Amin (1983). “Estimating parameters in continuous univariate distributions with a shifted origin”. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 394–403.

Chou, Yuan K (2002). “Testing alternative models of labour supply: Evidence from taxi drivers in Singapore”. *The Singapore Economic Review* 47.01, pp. 17–47.

Churchland, Anne K, Roozbeh Kiani, and Michael N Shadlen (2008). “Decision-making with multiple alternatives”. *Nature neuroscience* 11.6, pp. 693–702.

Cisek, Paul, Geneviève Aude Puskas, and Stephany El-Murr (2009). “Decisions in changing conditions: the urgency-gating model”. *The Journal of Neuroscience* 29.37, pp. 11560–11571.

Clithero, John A and Antonio Rangel (2013). *Combining response times and choice data using a neuroeconomic model of the decision process improves out-of-sample predictions*.

Corgnet, Brice, Roberto Hernán-González, and Eric Schniter (2015). “Why real leisure really matters: Incentive effects on real effort in the laboratory”. *Experimental Economics* 18.2, pp. 284–301.

Cox, David Roxbee and Hilton David Miller (1977). *The theory of stochastic processes*. Vol. 134. CRC Press.

Crawford, Vincent P and Juanjuan Meng (2011). "New york city cab drivers' labor supply revisited: Reference-dependent preferences with rationalexpectations targets for hours and income". *The American Economic Review* 101.5, pp. 1912–1932.

De Borger, Bruno and Mogens Fosgerau (2008). "The trade-off between money and travel time: A test of the theory of reference-dependent preferences". *Journal of Urban Economics* 64.1, pp. 101–115.

Deneve, Sophie (2012). "Making decisions with unknown sensory reliability". *Frontiers in neuroscience* 6.

Ding, Long and Joshua I Gold (2010). "Caudate encodes multiple computations for perceptual decisions". *The Journal of Neuroscience* 30.47, pp. 15747–15759.

– (2012). "Separate, causal roles of the caudate in saccadic choice and execution in a perceptual decision task". *Neuron* 75.5, pp. 865–874.

Ditterich, Jochen (2006a). "Evidence for time-variant decision making". *European Journal of Neuroscience* 24.12, pp. 3628–3641.

– (2006b). "Stochastic models of decisions about motion direction: behavior and physiology". *Neural Networks* 19.8, pp. 981–1012.

Doran, Kirk (2014). "Are long-term wage elasticities of labor supply more negative than short-term ones?" *Economics Letters* 122.2, pp. 208–210.

Douglas, RM et al. (2006). "Perception of visual motion coherence by rats and mice". *Vision research* 46.18, pp. 2842–2847.

Drugowitsch, Jan, Gregory C DeAngelis, Dora E Angelaki, et al. (2015). "Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making". *eLife* 4, e06678.

Drugowitsch, Jan, Gregory C DeAngelis, Eliana M Klier, et al. (2014). "Optimal multisensory decision-making in a reaction-time task". *Elife* 3, e03005.

Drugowitsch, Jan, Rubén Moreno-Bote, et al. (2012). "The cost of accumulating evidence in perceptual decision making". *The Journal of Neuroscience* 32.11, pp. 3612–3628.

Ebbinghaus, Hermann (1885). *Memory: A contribution to experimental psychology*.

Edwards, Ward (1965). "Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing". *Journal of Mathematical Psychology* 2.2, pp. 312–329.

Ekström, Magnus (2008). "Alternatives to maximum likelihood estimation based on spacings and the Kullback–Leibler divergence". *Journal of Statistical Planning and Inference* 138.6, pp. 1778–1791.

– (2013). "Powerful Parametric Tests Based on Sum-Functions of Spacings". *Scandinavian Journal of Statistics* 40.4, pp. 886–898.

Eliaz, Kfir and Rani Spiegler (2013). *Reference-dependence and labor-market fluctuations*. Tech. rep. National Bureau of Economic Research.

Ericson, Keith M Marzilli and Andreas Fuster (2011). "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments". *The Quarterly Journal of Economics* 126.4, pp. 1879–1907.

Farber, Henry S (2005). "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers". *Journal of Political Economy* 113.1, pp. 46–82.

– (2008). "Reference-dependent preferences and labor supply: The case of New York City taxi drivers". *The American Economic Review* 98.3, pp. 1069–1082.

– (2015). "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers". *The Quarterly Journal of Economics*.

Fehr, Ernst and Lorenz Goette (2007). "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment". *The American Economic Review*, pp. 298–317.

Feller, William (1968). *An introduction to probability theory and its applications: volume I*. Vol. 3. John Wiley & Sons London-New York-Sydney-Toronto.

Flôres Jr, Renato G and Victor A Ginsburgh (1996). "The Queen Elisabeth musical competition: how fair is the final ranking?" *The Statistician*, pp. 97–104.

Forstmann, Birte U, Alfred Anwander, et al. (2010). "Cortico-striatal connections predict control over speed and accuracy in perceptual decision making". *Proceedings of the National Academy of Sciences* 107.36, pp. 15916–15920.

Forstmann, Birte U, Gilles Dutilh, et al. (2008). "Striatum and pre-SMA facilitate decision-making under time pressure". *Proceedings of the National Academy of Sciences* 105.45, pp. 17538–17542.

Frazier, Peter and Angela J Yu (2008). "Sequential hypothesis testing under stochastic deadlines". *Advances in neural information processing systems*, pp. 465–472.

Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki (2015). "Stochastic Choice and Optimal Sequential Sampling". *Available at SSRN 2602927*.

Gill, David and Victoria Prowse (2012). "A structural analysis of disappointment aversion in a real effort competition". *The American Economic Review*, pp. 469–503.

Ginsburgh, Victor A and Jan C Van Ours (2003). "Expert opinion and compensation: Evidence from a musical competition". *The American Economic Review* 93.1, pp. 289–296.

Glejser, Herbert and Bruno Heyndels (2001). "Efficiency and inefficiency in the ranking in competitions: The case of the Queen Elisabeth Music Contest". *Journal of cultural Economics* 25.2, pp. 109–129.

Glenberg, Arthur M, Margaret M Bradley, Thomas A Kraus, et al. (1983). “Studies of the long-term recency effect: Support for a contextually guided retrieval hypothesis.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9.2, p. 231.

Glenberg, Arthur M, Margaret M Bradley, Jennifer A Stevenson, et al. (1980). “A two-process account of long-term serial position effects.” *Journal of Experimental Psychology: Human Learning and Memory* 6.4, p. 355.

Glenberg, Arthur M and Thomas A Kraus (1981). “Long-term recency is not found on a recognition test.” *Journal of Experimental Psychology: Human Learning and Memory* 7.6, p. 475.

Gold, Joshua I and Michael N Shadlen (2002). “Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward”. *Neuron* 36.2, pp. 299–308.

– (2007). “The neural basis of decision making”. *Annu. Rev. Neurosci.* 30, pp. 535–574.

Gul, Faruk (1991). “A theory of disappointment aversion”. *Econometrica*, pp. 667–686.

Guo, Pengfei and Paul Zipkin (2007). “Analysis and comparison of queues with different levels of delay information”. *Management Science* 53.6, pp. 962–970.

Hanes, Doug P and Jeffrey D Schall (1996). “Neural control of voluntary movement initiation”. *Science* 274.5286, pp. 427–430.

Hanks, Timothy D et al. (2011). “Elapsed decision time affects the weighting of prior probability in a perceptual decision task”. *The Journal of Neuroscience* 31.17, pp. 6339–6352.

Hassin, Refael (1986). “Consumer information in markets with random product quality: The case of queues and balking”. *Econometrica*, pp. 1185–1195.

Hawkins, Guy E et al. (2015). “Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making”. *The Journal of Neuroscience* 35.6, pp. 2476–2484.

Heath, Richard A (1982). “A model for signal detection based on an adaptive filter”. *Biological Cybernetics* 45.2, pp. 95–100.

Heathcote, Andrew, Scott Brown, and Denis Cousineau (2004). “QMPE: Estimating Lognormal, Wald, and Weibull RT distributions with a parameter-dependent lower bound”. *Behavior Research Methods, Instruments, & Computers* 36.2, pp. 277–290.

Heathcote, Andrew, Scott Brown, and D JK Mewhort (2002). “Quantile maximum likelihood estimation of response time distributions”. *Psychonomic bulletin & review* 9.2, pp. 394–401.

Heffetz, Ori and John A List (2014). “Is the endowment effect an expectations effect?” *Journal of the European Economic Association* 12.5, pp. 1396–1422.

Hockley, William E and Bennet B Murdock (1987). “A decision model for accuracy and response latency in recognition memory.” *Psychological Review* 94.3, p. 341.

Howard, Marc W and Michael J Kahana (1999). “Contextual variability and serial position effects in free recall.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25.4, p. 923.

Kahana, Michael Jacob (2012). *Foundations of human memory*. Oxford University Press.

Kardes, Frank R and Paul M Herr (1990). “Order effects in consumer judgment, choice, and memory: The role of initial processing goals”. *NA-Advances in Consumer Research Volume 17*.

Karşılar, Hakan et al. (2014). “Speed accuracy trade-off under response deadlines”. *Frontiers in neuroscience* 8, p. 248.

Kass, Robert E and Adrian E Raftery (1995). “Bayes factors”. *Journal of the american statistical association* 90.430, pp. 773–795.

Kim, Jong-Nam and Michael N Shadlen (1999). “Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque”. *Nature neuroscience* 2.2, pp. 176–185.

Kőszegi, Botond and Matthew Rabin (2006). “A model of reference-dependent preferences”. *The Quarterly Journal of Economics* 121.4, pp. 1133–1165.

Krajbich, Ian, Carrie Armel, and Antonio Rangel (2010). “Visual fixations and the computation and comparison of value in simple choice”. *Nature neuroscience* 13.10, pp. 1292–1298.

Krajbich, Ian, Dingchao Lu, et al. (2012). “The attentional drift-diffusion model extends to simple purchasing decisions”. *Frontiers in Psychology* 3.

Krajbich, Ian, Bastiaan Oud, and Ernst Fehr (2014). “Benefits of neuroeconomic modeling: New policy interventions and predictors of preference”. *The American Economic Review* 104.5, pp. 501–506.

Krajbich, Ian and Antonio Rangel (2011). “Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions”. *Proceedings of the National Academy of Sciences* 108.33, pp. 13852–13857.

LaBerge, David (1962). “A recruitment theory of simple behavior”. *Psychometrika* 27.4, pp. 375–396.

Lam, Edmund (2014). “The psychophysics of decision making in a two-direction random dot motion target selection task”.

Laming, Donald R J (1968). *Information theory of choice-reaction times.* Academic Press.

Leah-Martin, Vincent (2015). “How Much to Work: Evidence of Reference Dependence in Taxi Drivers”. *Available at SSRN 2646157*.

Li, Ye and Nicholas Epley (2009). “When the best appears to be saved for last: Serial position effects on choice”. *Journal of Behavioral Decision Making* 22.4, pp. 378–389.

Loomes, Graham and Robert Sugden (1986). “Disappointment and dynamic consistency in choice under uncertainty”. *The Review of Economic Studies*, pp. 271–282.

Mantonakis, Antonia et al. (2009). “Order in choice effects of serial position on preferences”. *Psychological Science* 20.11, pp. 1309–1312.

Markle, Alex et al. (2014). “Goals as reference points in marathon running: A novel test of reference dependence”. *Available at SSRN 2523510*.

Menéndez, M et al. (2001). “Minimum divergence estimators based on grouped data”. *Annals of the Institute of Statistical Mathematics* 53.2, pp. 277–288.

Meng, Juanjuan (2014). “Can prospect theory explain the disposition effect? A new perspective on reference points”. *Available at SSRN 1851883*.

Meredith, Marc and Yuval Salant (2013). “On the causes and consequences of ballot order effects”. *Political Behavior* 35.1, pp. 175–197.

Milosavljevic, Milica et al. (2010). “The Drift Diffusion Model Can Account for the Accuracy and Reaction Time of Value-based Choices under High and Low Time Pressure”. *Judgment and Decision Making* 5.6, pp. 437–449.

Mogilner, Cassie, Baba Shiv, and Sheena S Iyengar (2013). “Eternal quest for the best: Sequential (vs. simultaneous) option presentation undermines choice commitment”. *Journal of Consumer Research* 39.6, pp. 1300–1312.

Moran, Rani (2015). “Optimal decision making in heterogeneous and biased environments”. *Psychonomic bulletin & review* 22.1, pp. 38–53.

Murdock, Bennet and Janet Metcalfe (1978). “Controlled rehearsal in single-trial free recall”. *Journal of Verbal Learning and Verbal Behavior* 17.3, pp. 309–324.

Neath, Ian (1993). “Contextual and distinctive processes and the serial position function”. *Journal of Memory and Language* 32.6, p. 820.

Newsome, William T, Kenneth H Britten, and J Anthony Movshon (1989). “Neuronal correlates of a perceptual decision”. *Nature* 341.6237, pp. 52–54.

Nguyen, Angela P et al. (2004). “A dissociation of motion and spatial-pattern vision in the avian telencephalon: implications for the evolution of ?visual streams?” *The Journal of Neuroscience* 24.21, pp. 4962–4970.

O'Connell, Redmond G, Paul M Dockree, and Simon P Kelly (2012). "A supramodal accumulation-to-bound signal that determines perceptual decisions in humans". *Nature neuroscience* 15.12, pp. 1729–1735.

Page, Lionel and Katie Page (2010). "Last shall be first: A field study of biases in sequential performance evaluation on the Idol series". *Journal of economic behavior & organization* 73.2, pp. 186–198.

Pagel, Michaela (2013). "Expectations-based reference-dependent life-cycle consumption". *Available at SSRN 2268254*.

Payne, John W et al. (2000). "Valuation of multiple environmental programs". *Journal of Risk and Uncertainty* 21.1, pp. 95–115.

Philiastides, Marios G, Roger Ratcliff, and Paul Sajda (2006). "Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram". *The Journal of neuroscience* 26.35, pp. 8965–8975.

Pike, AR (1966). "Stochastic models of choice behaviour: response probabilities and latencies of finite Markov chain systems." *The British journal of mathematical and statistical psychology* 19.1, pp. 15–32.

Poltrock, Steven E and Colin M MacLeod (1977). "Primacy and recency in the continuous distractor paradigm." *Journal of Experimental Psychology: Human Learning and Memory* 3.5, p. 560.

Pope, Devin G and Maurice E Schweitzer (2011). "Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes". *The American Economic Review* 101.1, pp. 129–157.

Raftery, Adrian E (1995). "Bayesian model selection in social research". *Sociological methodology*, pp. 111–163.

Ramakrishnan, Arjun and Aditya Murthy (2013). "Brain mechanisms controlling decision making and motor planning". *Prog Brain Res* 202, pp. 321–345.

Ramakrishnan, Arjun, Ramakrishnan Sureshbabu, and Aditya Murthy (2012). "Understanding how the brain changes its mind: microstimulation in the macaque frontal eye field reveals how saccade plans are changed". *The Journal of Neuroscience* 32.13, pp. 4457–4472.

Ranneby, Bo (1984). "The maximum spacing method. An estimation method related to the maximum likelihood method". *Scandinavian Journal of Statistics*, pp. 93–112.

Rao, Rajesh PN (2010). "Decision making under uncertainty: a neural model based on partially observable markov decision processes". *Frontiers in computational neuroscience* 4, p. 146.

Ratcliff, Roger (1978). "A theory of memory retrieval." *Psychological review* 85.2, p. 59.

Ratcliff, Roger (1980). "A note on modeling accumulation of information when the rate of accumulation changes over time". *Journal of Mathematical Psychology* 21.2, pp. 178–184.

– (2002). "A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data". *Psychonomic bulletin & review* 9.2, pp. 278–291.

Ratcliff, Roger, Anil Cherian, and Mark Segraves (2003). "A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions". *Journal of neurophysiology* 90.3, pp. 1392–1407.

Ratcliff, Roger, Marios G Philiastides, and Paul Sajda (2009). "Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG". *Proceedings of the National Academy of Sciences* 106.16, pp. 6539–6544.

Ratcliff, Roger and Jeffrey N Rouder (1998). "Modeling response times for two-choice decisions". *Psychological Science* 9.5, pp. 347–356.

– (2000). "A diffusion model account of masking in two-choice letter identification." *Journal of Experimental Psychology: Human perception and performance* 26.1, p. 127.

Ratcliff, Roger and Philip L Smith (2004). "A comparison of sequential sampling models for two-choice reaction time". *Psychological review* 111.2, p. 333.

Robertson, Tim (1978). "Testing for and against an order restriction on multinomial parameters". *Journal of the American Statistical Association* 73.361, pp. 197–202.

Robertson, Tim, F.T. Wright, and Richard Dykstra (1988). *Order restricted statistical inference*.

Rotthoff, Kurt W (2015). "(Not Finding a) Sequential Order Bias in Elite Level Gymnastics". *Southern Economic Journal* 81.3, pp. 724–741.

Rubinstein, Ariel (2007). "Instinctive and Cognitive Reasoning: A Study of Response Times". *The Economic Journal* 117.523, pp. 1243–1259.

Rundus, Dewey (1971). "Analysis of rehearsal processes in free recall." *Journal of experimental psychology* 89.1, p. 63.

Salant, Yuval (2011). "Procedural analysis of choice rules with applications to bounded rationality". *The American Economic Review* 101.2, pp. 724–748.

Salinas, Emilio and Terrence R Stanford (2013). "The countermanding task revisited: fast stimulus detection is a key determinant of psychophysical performance". *The Journal of Neuroscience* 33.13, pp. 5668–5685.

Sanders, AF and W Ter Linden (1967). "Decision making during paced arrival of probabilistic information". *Acta psychologica* 27, pp. 170–177.

Schall, Jeffrey D (2003). “Neural correlates of decision processes: neural and mental chronometry”. *Current opinion in neurobiology* 13.2, pp. 182–186.

Schosser, Stephan, Judith N Trarbach, and Bodo Vogt (2013). “Is the first option worth paying for? An experimental study on the primacy effect”. *Diskussionsbeitrag Nr. 36-13*.

Schurger, Aaron, Jacobo D Sitt, and Stanislas Dehaene (2012). “An accumulator model for spontaneous neural activity prior to self-initiated movement”. *Proceedings of the National Academy of Sciences* 109.42, E2904–E2913.

Shadlen, Michael N and William T Newsome (1996). “Motion perception: seeing and deciding”. *Proceedings of the National Academy of Sciences* 93.2, pp. 628–633.

– (2001). “Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey”. *Journal of neurophysiology* 86.4, pp. 1916–1936.

Shiryaev, AN (1969). “Sequential statistical analysis”. *Translations of Mathematical Monographs* 38.

Simen, Patrick et al. (2009). “Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions.” *Journal of Experimental Psychology: Human Perception and Performance* 35.6, p. 1865.

Smith, Alec (2012). “Lagged Beliefs and Reference-Dependent Preferences”.

Smith, Philip L and Cameron RL McKenzie (2011). “Diffusive information accumulation by minimal recurrent neural models of decision making”. *Neural computation* 23.8, pp. 2000–2031.

Smith, Philip L and Roger Ratcliff (2004). “Psychology and neurobiology of simple decisions”. *Trends in neurosciences* 27.3, pp. 161–168.

Smith, Philip L, Roger Ratcliff, and Bradley J Wolfgang (2004). “Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays”. *Vision research* 44.12, pp. 1297–1320.

Spiliopoulos, Leonidas and Andreas Ortmann (2014). “The BCD of Response Time Analysis in Experimental Economics”. *Available at SSRN 2401325*.

Starns, Jeffrey J and Roger Ratcliff (2010). “The effects of aging on the speed–accuracy compromise: Boundary optimality in the diffusion model.” *Psychology and aging* 25.2, p. 377.

– (2012). “Age-related differences in diffusion model boundary optimality with both trial-limited and time-limited tasks”. *Psychonomic bulletin & review* 19.1, pp. 139–145.

– (2014). “Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis”. *Journal of memory and language* 70, pp. 36–52.

Stone, Mervyn (1960). “Models for choice-reaction time”. *Psychometrika* 25.3, pp. 251–260.

Talmi, Deborah and Yonatan Goshen-Gottstein (2006). “The long-term recency effect in recognition memory”. *Memory* 14.4, pp. 424–436.

Tan, Lydia and Geoff Ward (2000). “A recency-based account of the primacy effect in free recall.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26.6, p. 1589.

Thura, David, Julie Beauregard-Racine, et al. (2012). “Decision making by urgency gating: theory and experimental support”. *Journal of neurophysiology* 108.11, pp. 2912–2930.

Thura, David and Paul Cisek (2014). “Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making”. *Neuron* 81.6, pp. 1401–1416.

Tuerlinckx, Francis et al. (2001). “A comparison of four methods for simulating the diffusion process”. *Behavior Research Methods, Instruments, & Computers* 33.4, pp. 443–456.

Usher, Marius and James L McClelland (2001). “The time course of perceptual choice: the leaky, competing accumulator model.” *Psychological review* 108.3, p. 550.

Vickers, D (1970). “Evidence for an accumulator model of psychophysical discrimination”. *Ergonomics* 13.1, pp. 37–58.

Viviani, P and CA Terzuolo (1972). “On the modeling of the performances of the human brain in a two-choice task involving decoding and memorization of simple visual patterns”. *Kybernetik* 10.3, pp. 121–137.

Viviani, Paolo (1979a). “A diffusion model for discrimination of temporal numerosity”. *Journal of Mathematical Psychology* 19.2, pp. 108–136.

– (1979b). “Choice reaction times for temporal numerosity.” *Journal of experimental psychology: human perception and performance* 5.1, p. 157.

Voskuilen, Chelsea, Roger Ratcliff, and Philip L Smith (2016). “Comparing fixed and collapsing boundary versions of the diffusion model”. *Journal of Mathematical Psychology* 73, pp. 59–79.

Voss, Andreas, Klaus Rothermund, and Jochen Voss (2004). “Interpreting the parameters of the diffusion model: An empirical validation”. *Memory & Cognition* 32.7, pp. 1206–1220.

Wagenmakers, Eric-Jan, Han LJ Van Der Maas, and Raoul PPP Grasman (2007). “An EZ-diffusion model for response time and accuracy”. *Psychonomic bulletin & review* 14.1, pp. 3–22.

Wagner, Judy A and Noreen M Klein (2007). "Who wants to go first? order effects within a series of competitive sales presentations". *Journal of Personal Selling & Sales Management* 27.3, pp. 259–276.

Wald, A (1947). *Sequential analysis. 1947*.

Wald, Abraham and Jacob Wolfowitz (1948). "Optimum character of the sequential probability ratio test". *The Annals of Mathematical Statistics* 19.3, pp. 326–339.

Wang, Xiao-Jing (2002). "Probabilistic decision making by slow reverberation in cortical circuits". *Neuron* 36.5, pp. 955–968.

Ward, Geoff (2002). "A recency-based account of the list length effect in free recall". *Memory & Cognition* 30.6, pp. 885–892.

Wasserman, Larry (2000). "Bayesian model selection and model averaging". *Journal of mathematical psychology* 44.1, pp. 92–107.

Watkins, Michael J, Ian Neath, and Elizabeth S Sechler (1989). "Recency effect in recall of a word list when an immediate memory task is performed after each word presentation". *The American journal of psychology* 102.2, pp. 265–270.

Webb, Ryan (2013). "Dynamic Constraints on the Distribution of Stochastic Choice: Drift Diffusion Implies Random Utility". *Available at SSRN 2226018*.

Wilks, Samuel S (1938). "The large-sample distribution of the likelihood ratio for testing composite hypotheses". *The Annals of Mathematical Statistics* 9.1, pp. 60–62.

Wilson, Vietta E (1977). "Objectivity and effect of order of appearance in judging of synchronized swimming meets". *Perceptual and Motor Skills* 44.1, pp. 295–298.

Wimmer, G Elliott, Erin Kendall Braun, et al. (2014). "Episodic memory encoding interferes with reward learning and decreases striatal prediction errors". *The Journal of Neuroscience* 34.45, pp. 14901–14912.

Wimmer, G Elliott and Daphna Shohamy (2012). "Preference by association: how memory mechanisms in the hippocampus bias decisions". *Science* 338.6104, pp. 270–273.

Wong, Kong-Fatt and Xiao-Jing Wang (2006). "A recurrent network mechanism of time integration in perceptual decisions". *The Journal of neuroscience* 26.4, pp. 1314–1328.

Woodford, Michael (2014). "Stochastic choice: An optimizing neuroeconomic model". *The American Economic Review* 104.5, pp. 495–500.

Yang, Liu, Pengfei Guo, and Yulan Wang (2014). "Service pricing with loss averse customers". *Available at SSRN 2418303*.

Yu, Qiuping, Gad Allon, and Achal Bassamboo (forthcoming). "How do delay announcements shape customer behavior? An empirical study". *Management Science*.

Zacksenhouse, M, R Bogacz, and P Holmes (2010). "Robust versus optimal strategies for two-alternative forced choice tasks". *Journal of mathematical psychology* 54.2, pp. 230–246.

Zhang, Shunan et al. (2014). "Time-varying boundaries for diffusion models of decision making and response time". *Frontiers in psychology* 5.

Zohar, Ety, Avishai Mandelbaum, and Nahum Shimkin (2002). "Adaptive behavior of impatient customers in tele-queues: Theory and empirical support". *Management Science* 48.4, pp. 566–583.

Appendix A

EXPERIMENTAL INSTRUCTIONS FOR CHAPTER 2

Experienced condition:

‘Welcome to this experiment in perceptual decision making. It consists of two blocks with a short break in between. When you complete the experiment, please remain seated until half an hour has passed from the beginning of the experiment. You may browse the internet in the meantime. After 35 minutes the experiment will automatically conclude. Please slide out the partitions beside you. Adjust the height and position of your chair so that your eye level is at two-thirds of the screen height. When you are ready, hit SPACE to continue.’

‘This part of the experiment involves categorizing dot motion. It consists of 100 trials. In each trial, many dots will appear on the screen, moving in different directions. Some of the dots are signal dots moving in a consistent direction. Noise dots with random motion are overlaid on top of the signal dots. The task is to choose the signal dot direction. If you think they are moving to the left, hit “1”. If you think they are moving to the right, hit “2”. Two examples will be displayed on the next screens. At the end of the experiment you will receive $0.05 for each correct answer and nothing for each incorrect answer. When you are ready, hit SPACE to continue.’

‘The following two trials do not count toward your final totals. A left trial will appear first, and then a right trial. When you are ready, hit SPACE to continue.’

‘Some of the dots are signal dots moving in a consistent direction. The task is to choose this direction. Hit “1” or “2” to continue.’

‘Noise dots with random motion are overlaid on top of the signal dots. This is what you will see. The task is to choose the signal dot direction. Hit “1” or “2” to continue.’

‘Tip: Try not to focus on individual dots. Instead, look for spread out sets of dots that are moving rigidly together, and follow them as a whole. The set of dots may emerge, or "pop out" from the rest. When you are ready, hit SPACE to continue.’

‘The real trials begin now. When you are ready, hit SPACE to continue.’

(Block 1)

'Take a one minute break. After a minute, instructions for the next part will automatically appear.'

'This is the final part, and should take about 10 minutes to complete. It is the same as the previous part, categorizing dot motion. When you are ready, hit SPACE to continue.'

(Block 2)

'Before we finish, please answer the following few questions. Hit SPACE to continue.'

'On a scale of 1 to 10, how much did you like the task, with 1 being very little and 10 being very much? Use your mouse to answer.'

'What percentage of answers do you think you got correct in the first task block? (Bear in mind that random choices would lead to 50% accuracy.) Use your mouse to answer.'

'What percentage of answers do you think you got correct in the second task block? (Bear in mind that random choices would lead to 50% accuracy.) Use your mouse to answer.'

'Have you ever participated in this kind of task before in another experiment? Use your mouse to answer. (0 = No, 1 = Yes)'

'Thanks for your participation. Remember to wait in your seat until directed otherwise. You may browse the internet in the meantime. Hit SPACE to finish.'

Inexperienced condition:

'Welcome to this experiment in perceptual decision making. It consists of two blocks with a short break in between. When you complete the experiment, please remain seated until half an hour has passed from the beginning of the experiment. You may browse the internet in the meantime. After 35 minutes the experiment will automatically conclude. Please slide out the partitions beside you. Adjust the height and position of your chair so that your eye level is at two-thirds of the screen height. When you are ready, hit SPACE to continue.'

‘This part of the experiment involves categorizing blurred images. It consists of 100 trials. Each image contains either a raccoon or a porcupine. If you think it is a raccoon, hit “1”. If you think it is a porcupine, hit “2”. At the end of the experiment you will receive $0.05 for each correct answer and nothing for each incorrect answer. Examples will be displayed on the next screens. When you are ready, hit SPACE to continue.’

‘The following two trials do not count toward your final totals. A raccoon image will appear first, and then a porcupine image. The real trials will be blurred and grayed images. Hit either “1” or “2” to end each trial. When you are ready, hit SPACE to continue.’

‘The real trials begin now. When you are ready, hit SPACE to continue.’

(Block 1)

‘Take a one minute break. After a minute, instructions for the next part will automatically appear.’

‘This is the final part, and should take about 10 minutes to complete. It involves categorizing dot motion, and consists of 100 trials. In each trial, many dots will appear on the screen, moving in different directions. Some of the dots are signal dots moving in a consistent direction. Noise dots with random motion are overlaid on top of the signal dots. The task is to choose the signal dot direction. If you think they are moving to the left, hit “1”. If you think they are moving to the right, hit “2”. Two examples will be displayed on the next screens. At the end of the experiment you will receive $0.05 for each correct answer and nothing for each incorrect answer. When you are ready, hit SPACE to continue.’

‘The following two trials do not count toward your final totals. A left trial will appear first, and then a right trial. When you are ready, hit SPACE to continue.’

‘Some of the dots are signal dots moving in a consistent direction. The task is to choose this direction. Hit “1” or “2” to continue.’

‘Noise dots with random motion are overlaid on top of the signal dots. This is what you will see. The task is to choose the signal dot direction. Hit “1” or “2” to continue.’

‘Tip: Try not to focus on individual dots. Instead, look for spread out sets of dots that are moving rigidly together, and follow them as a whole. The set of dots may emerge, or "pop out" from the rest. When you are ready, hit SPACE to continue.’

‘The real trials begin now. When you are ready, hit SPACE to continue.’

(Block 2)

‘Before we finish, please answer the following few questions. Hit SPACE to continue.’

‘On a scale of 1 to 10, how much did you like the task, with 1 being very little and 10 being very much? Use your mouse to answer.’

‘What percentage of answers do you think you got correct in the first task block? (Bear in mind that random choices would lead to 50% accuracy.) Use your mouse to answer.’

‘What percentage of answers do you think you got correct in the second task block? (Bear in mind that random choices would lead to 50% accuracy.) Use your mouse to answer.’

‘Have you ever participated in this kind of task before in another experiment? Use your mouse to answer. (0 = No, 1 = Yes)’

‘Thanks for your participation. Remember to wait in your seat until directed otherwise. You may browse the internet in the meantime. Hit SPACE to finish.’

Appendix B

MODEL INVESTIGATIONS FOR CHAPTER 2

B.1 Accuracy comparisons

Figure B.1.1 illustrates the effect of different accuracy functions on time choice. The effect of shifting accuracies is ambiguous. Roughly speaking, a "more accurate" person will spend less time than someone who is less accurate, and conclusions about overall accuracy cannot be drawn (Fig 2a), unless there is time pressure from the reference point, in which case the reverse may hold (Fig 2b). This applies even when "more accurate" entails a strictly greater accuracy for any given time expenditure. This occurs because under time pressure, the less accurate person may be unable to produce accuracy quickly enough, so to speak, to justify spending as much time in the attempt. What determines whether or not there is time pressure is the crossing point between the marginal accuracies – the time at which the marginal accuracies are equal. Under the assumption that each person has the same baseline accuracy $a(0)$ and asymptotic accuracy $\lim_{t\to\infty} a(t)$ as everyone else, there always exists at least one such point for each pair of accuracy functions,[1] though it is not generally the same point in different pairs. If there are fundamentally different capabilities in baseline or asymptotic terms, then a crossing point doesn't necessarily exist,[2] and it is possible to have improvements in accuracy capability which lead to an unambiguous weak increase in time spent and consequently in accuracy; the time pressure result holds without time pressure (Fig 2c).

[1]Suppose $a_h(t)$ and $a_l(t)$ are two accuracy functions such that there is no t for which $a'_h(t) = a'_l(t)$, so wlog $a'_h(t) > a'_l(t)$. Then by the fundamental theorem of calculus, $\lim_{t\to\infty} a_h(t) - a_h(0) = \int_0^\infty a'_h(t)dt > \int_0^\infty a'_l(t)dt = \lim_{t\to\infty} a_l(t) - a_l(0)$, contradicting the assumption.

[2]Having no crossing point requires $\lim_{t\to\infty} a_h(t) - a_h(0) > \lim_{t\to\infty} a_l(t) - a_l(0)$. If everyone asymptotes at the same accuracy, then $a_h(0) < a_l(0)$ is necessary for no crossing point; the higher marginal accuracy individual must have worse baseline accuracy. If everyone begins with the same accuracy, then $\lim_{t\to\infty} a_h(t) > \lim_{t\to\infty} a_l(t)$ is necessary for no crossing point; the higher marginal accuracy individual must have better asymptotic accuracy.

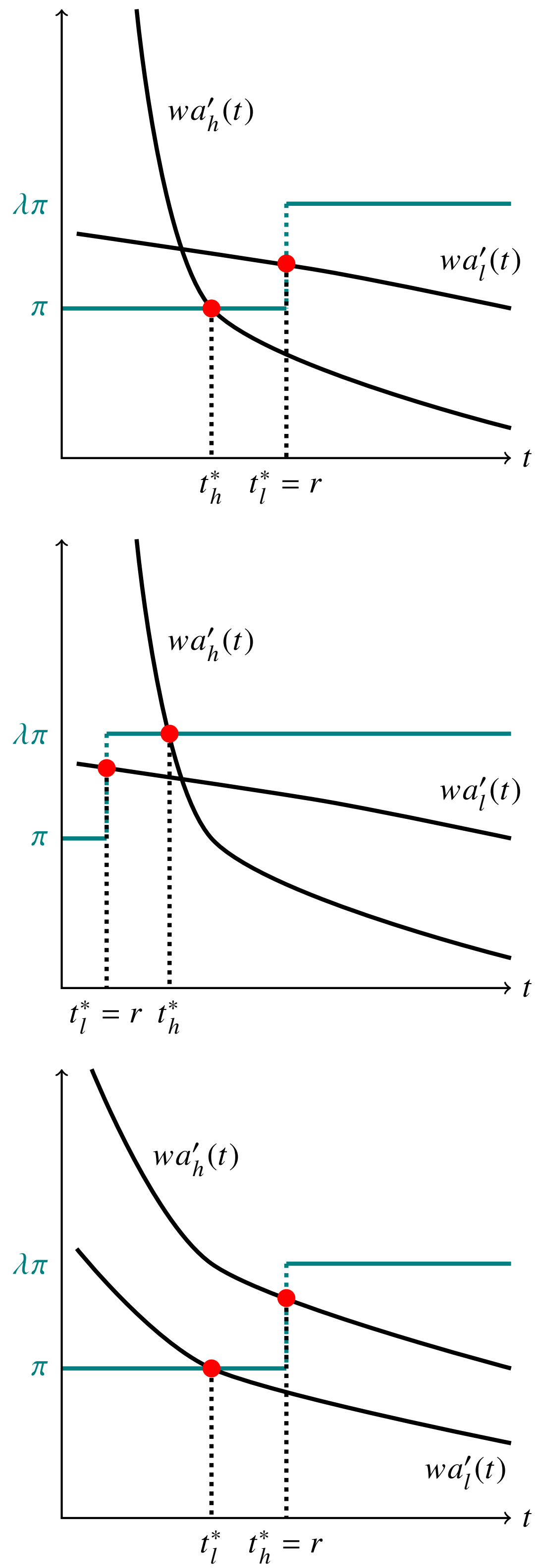

Figure B.1.1: Comparison of behavior across different accuracy functions

B.2 Dynamic modeling

To illustrate the behavioral dynamics of a farsighted agent, I first write out a simple version of a dynamic model. The reference-dependent utility function is

$$U(t|r) = wa(t) + \begin{cases} -\pi t & \text{if } t < r \\ -\pi r - \lambda\pi(t-r) & \text{if } t \geq r \end{cases}$$

where λ is the coefficient of loss aversion. The marginal utilities in each case are:

$$t < r : \frac{\partial U}{\partial t} = wa'(t) - \pi$$
$$t > r : \frac{\partial U}{\partial t} = wa'(t) - \lambda\pi$$

Three cases can occur that are distinguished by how quickly an agent with reference-dependent preferences would finish compared to their standard expected utility counterpart.[3] A standard economic agent chooses $\tilde{t}$ such that $wa'(\tilde{t}) = \pi$. If $\tilde{t} \leq r$, meaning that the individual would finish before the reference point is hit even in the absence of loss aversion, then the reference point does not threaten them at all and so they will make the same choice,[4] $t^* = \tilde{t}$ (Figure B.2.2a). If $\tilde{t} > r$ then the individual faces an added psychological cost when exceeding the reference point. They will curtail their time expenditure to avoid being subjectively penalized at a higher rate, choosing $t^* < \tilde{t}$. Within this regime, if $\tilde{t}$ is not too much greater than r, the individual will work right up until the reference point and stop due to the discontinuous jump in cost, so $t^* = r$ (Figure B.2.2b). If $\tilde{t}$ is significantly greater than r, they will continue to work even past the reference point, although still less than their standard counterpart (Figure B.2.2c).[5] Reference dependence restrains people from exceeding the reference point due to dramatically higher marginal costs. And since these people are spending less time than they otherwise would have, they become less accurate. The strength of these effects is tied to the severity of loss aversion.

---

[3] Appendix B contains preliminary investigations into accuracy function comparisons and choice dynamics.

[4] Although in this simple formulation there is no difference for those who would normally finish sufficiently quickly, introducing noise such as trembles in performance would induce time shading to reduce the risk of missing the reference point. This expands the prediction's range of effectiveness.

[5] To be more precise, let t_1 and t_2 be the points where the marginal benefit curve intersects the potential marginal cost curves such that $wa'(t_1) = \pi$ and $wa'(t_2) = \lambda\pi$. Clearly $t_1 > t_2$ given $\lambda > 1$. If the reference point is above t_1, then t_1 is chosen. If the reference point is between t_1 and t_2, then time is spent just up to the reference point. If the reference point is below t_2, then t_2 is chosen.

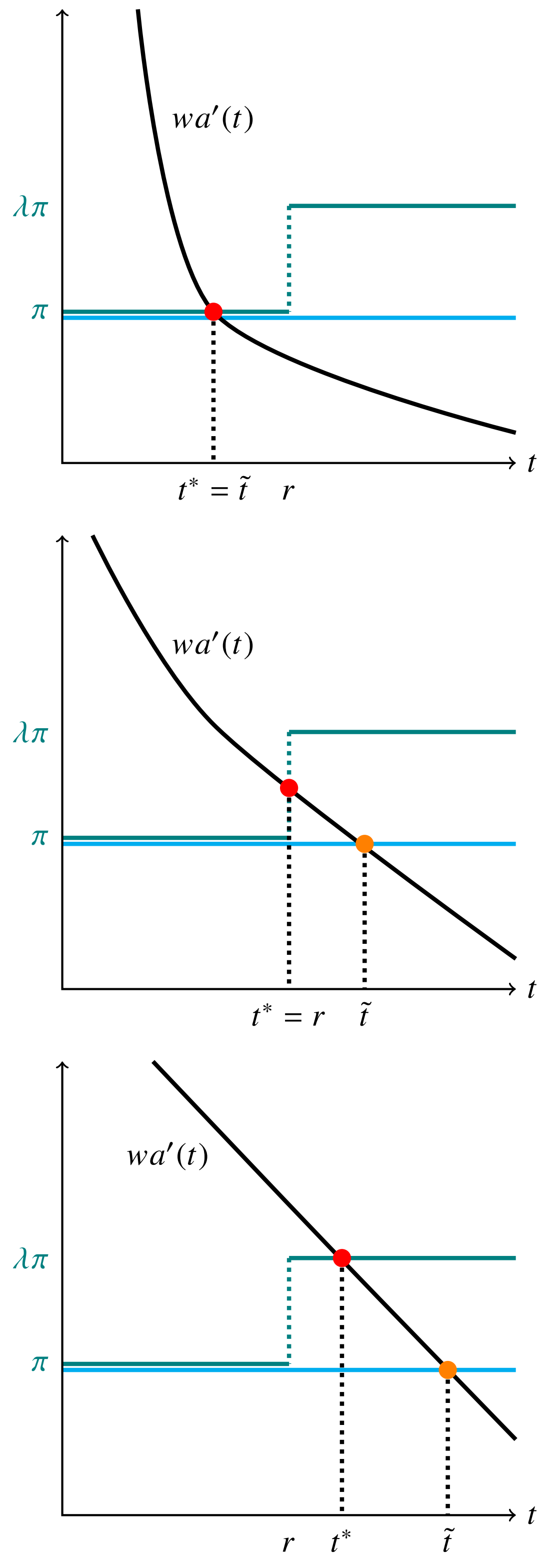

Figure B.2.2: Time choices with reference-dependent preferences

In the more complex dynamic version of this model, the agent maximizes his total expected payoff from the current trial through the remaining L trials. He may have different accuracy functions in every trial, which could occur due to learning, training, or boredom. He does not necessarily know what they will be, but has a belief over each of their distributions, though these beliefs may be mistaken. The expectation for each of these distributions gives rise to an accuracy function which has the desired properties; it is concave, increasing in t, and bounded. His expectation at trial j looking forward to trial ℓ's accuracy function is denoted $E_j[a_\ell(t)]$. It is supposed that $E_j[E_k[a_\ell(t)]] = E_j[a_\ell(t)]$ for $j \leq k \leq \ell$, which is a natural consistency requirement. If the agent believed he would change his mind, he should alter his beliefs to be consistent in the first place. He thus faces the following problem:

$$\max_{\{t_\ell\}_{\ell=1}^{L}} \sum_{\ell=1}^{L} wE_1[a_\ell(t_\ell)] + \begin{cases} -\pi t_\ell & \text{if } \sum_{\ell=1}^{L} t_\ell < R \\ -\pi R - \lambda\pi \left(\sum_{\ell=1}^{L} t_\ell - R\right) & \text{if } \sum_{\ell=1}^{L} t_\ell \geq R \end{cases}$$

As in the simpler version, the solution breaks down into three cases. In all cases, the agent chooses to allocate time such that expected marginal accuracies are equalized across all trials: $E_1[a'_\ell(t^*_\ell)] = E_1[a'_m(t^*_m)]\ \forall\, \ell, m$. In the first case, the optimal choice is $\{t^*_\ell\}$ such that $wE_1[a'_\ell(t^*_\ell)] = \pi\ \forall\, \ell$ as long as $\sum_{\ell=1}^{L} t^*_\ell < R$. In the second case, the optimal choice satisfies $wE_1[a'_\ell(t^*_\ell)] = \lambda\pi\ \forall \ell$ as long as $\sum_{\ell=1}^{L} t^*_\ell \geq R$. If the tentative values of t^*_ℓ are inconsistent with the reference point condition in both cases (that is, if the $\{t^*_\ell\}$ satisfying the equality does not satisfy the inequality), then $\sum_{\ell=1}^{L} t^*_\ell = R$ is the constraint which must be satisfied while equalizing marginal accuracies. Notice that if the accuracy function does not change, then the model effectively reduces to the one-period model with $r = R/L$.

Suppose the third case obtains, which reflects most clearly the impact of reference dependence. For similar reasons as when comparing different accuracy functions in the static setup, unambiguous predictions about overall accuracy in high versus low capability trials are unavailable. If, for illustration, we assume higher capability trials exhibit greater baseline accuracy, strictly higher marginal accuracy, and stronger concavity of the accuracy function (i.e. marginal accuracy curve is more strongly negatively sloped), an agent who anticipates improvement over the course of trials will spend less time and be accordingly less accurate on earlier trials.

We can say a bit more, though. Making the reference point a little tighter (i.e. shortening the time) leads the agent to cut back on time differently in each trial.

He cuts back more severely on trials in which $a'(t^*)$ is flatter, which is when the accuracy function diminishes at a slower rate. Generally speaking, this describes worse accuracy functions, which are less strongly concave. Thus when increasingly pressured, the agent disproportionately takes time away from trials on which his abilities are relatively poorer.

A tightening of the reference point is similar to what the agent faces when he makes a miscalculation and overestimates how much time he has left or how much he will improve by. Suppose he continually makes this misperception. In the simplest case, his accuracy function is stationary and he recognizes this. Then he will of course continually spend less and less time as he keeps realizing he would otherwise miss the reference point, and get less and less accurate as trials pass. Or, under the plethora of illustrative assumptions above, the agent will continually shave his current trial down from what he had been expecting before. This will attenuate the upward trend in time and accuracy.

Appendix C

SELF-REPORTED MEMORY STRATEGIES FOR CHAPTER 4

Response
I tried to associate a short word w/ each picture & repeated the words in order in my mind
I tried to remember a 1 word description for each image
I tried a few different thing: -Usually I tried to assign an object to each image that looked vaguely like the image -I tried to remember the colors -Sometimes I traced the image with my hands
I tried to think of names to associate with each image. The name was a color or shape or both
Tried assigning 1-word description to each image to remember later
-Attempt to associate images w/shape (i.e. any evident curves) color palette (i.e. light & dark contrast) texture (i.e. "lasers")
I associated each image with a word to reduce the problem to remembering a phrase.
I tried to assign each fractal a one- or two-word phrase based on its color or shape. I then repeated these phrases in order as I progressed through the round.
I tried to give the paintings titles and I placed a finger on each number as it appeared. Singing the alphabet also helps.
1st: Colors were associated with order 2nd: Then shapes 3rd: Finally, if the image reminded me of anything.
color - most prominent shape (rectangle, circle, lines)
I tried to remember empty spaces + shapes
Try to relate to real images. For instance -Looks like leaves -Looks like a sun & moon -Looks like building or -One half of it is solid colors

-Would focus on overall shape & not the details that could not be seen when blurred.
tried to remember colors and likening shapes to pictures I was already familiar with.
I tried to relate the pictures I see to lightening, circles, letters, the moon, etc
At the beginning: -For ones including letter exercise: Focused more on letters and not so much on pictures. -For "please wait" exercises: tried coming up with 1-word descriptions of pictures and making a sentence with them. By the end: -For ones including letter exercise: tried going through letters as quickly as possible so pictures stayed in my mind. -For "please wait" exercises: just focused on whether I liked the picture or not.
Associate each image with a word or words; then make a story consisting 5 images in order.
-Colors dominating each image -reference to remind of a scene or event
I tried to associate the picture with some familiar object
I tried to remember distinctive features of each photo.
I tried to remember some special shapes or color distribution in the images
Give the pictures a title or imagine seeing them as a figure, or character, or object
I tried to match descriptive words to each picture (e.g. scales, ripple, fire, Julia, bug (yay fractals!)). Remembering was hard though.
Tried to remember color scheme by relating it to an object (fire, beach, leaves, etc)
I tried to think of an object that resembled to image, either in shape or in color. Then I would be able to form a string of 5 words to try and remember the images & their order.
Generalized color and/or color scheme and shape e.g. yellow spiral, pink + grey, GREEN
For each image, come up with word/short phrase to help remember it later on eg fish, green-gold, circles, wave, etc and then keep the ordering of words in my mind so I know which number to click
I tried to keep repeating the main colors in my head
I associated a word with each image and repeated the words in sequence in my head.
First word in my head for each picture try to make a story

I tried remembering the color & a descriptive word about each image & tried to associate those words/color with the number of the image (1-5). Sometimes I'd remember less if the shape was super simple, or more if i felt it necessary. eg. "blue cross", "red crescent", "teal dots", &c.
Imagine scenario for the picture Imagine downloading picture into a finger (1-5) Imagine 5 pics around my hand to remember.
1-2 word description of what the images resembled, e.g. "yellow seahorse"
To begin I didn't understand the instructions fully and put 0 for everything. Once I did understand though I tried to remember a word the image reminded me of to recall the image and position
identify shapes/significant features/colors in each image
I tried to memorize by geometric representations of objects (triangles, circles, etc.) and also via symmetry. I tried using color as well, but I didn't think that worked too well.
I associated words of what the pictures looked liked, and repeated them in my head in order.
match the image with a word memorize the word
Using color, texture, shape Associating w/words + telling a story
I tried to remember a sequence of one or two word descriptions of the images, such as "lightning" or "green swirl". I recited them to myself as I progressed through each block, though the letters usually got in the way.
Came up w/ name/description to remind me of each [?]
For ones without letters between, I just tried to remember a single distinctive feature (large black blob or something) If there were letters in between, I tried to remember the last one and guessed '0' for everything else
tried to come up with short phrases to remember the pictures.
Tried to associate images with patterns/objects I saw to remember them.
1) remembered colors 2) remembered side w/largest concentration of stuff
-visualized my home as if art was hanging on wall -used names on fingers "shape of mandelbrot set"

-connected images to those of sexual ones (really) -imagined if paintings were hung in this room
I was unsure from the initial description whether we were allowed to memorize things like color and shape, but by the middle I did, trying to remember phrases like "blue and white CDs" or "red and white virus" to recall colors and defining forms.
First I tried associating each image with a word and to remember the word with the number of image it was. Then I tried remembering the color and pattern of each image.
-To remember the order of images I associated a word with each one of them and repeated the words in my head while doing the experiment. -Before each new section I repeated the alphabet in my head to help me remember the correct order of letters.
memorize color. Use word like lightning, sun, big bang, star, seahorse or so
Color, defining figures/shapes, ignored alphabet
I tried to remember a key word to remember the images, such as "leaf" or "neurons" or "firework". It was difficult because most of the images were fairly abstract so there was not something that they always remembered. I would think "fractal", but then 3 of the 5 would look like fractals so that became a useless description.
I would try to remember what the shapes reminded me of. If that didn't work, I would recognize the actual shape and then the color.
Associate pictures by color or object
1. Try to make story from seq of pics 2. keep finger on number keys to keep track of which index was shown, i.e. finger on 4 if this is 4th pic
I tried to remember colors/shapes I also tried to assign a word to pictures & tried to remember the word sequence.
I remember images using objects that fit the main features. e.g. snow [?], five flames, fractle, etc.
tried to associate words w/ images. This included themes, colors, etc.
assigned each picture a name, memorized all the names.
I either made up a name for each picture, or chose a dominant color for each to remember it by. I also kept my fingers on the number keys an touched the key when I thought the name/color.
attempted to come up w/ one word description and remember that

I tried to repeat color & general shape in my head in order (e.g. rainbow triangle-y shape, white blue spiral). Sometimes I also tried to add the numer to this short description to remember. A few rounds I tried to only look and not think too hard, just implant an impression of the image in my mind (this only worked when I didn't need to type letters in between).
I tried to give a title to each art piece that was descriptive but short. Alternatively, I would try to imagine the art as a physical 3D object.
repeating characteristics of the image "black star", "green fern", etc.
Memorizing the dominant shapes in the images and repeating them in my head
Strategy: (for images) -Find one or two words that characterize the dominating features of the image, such as "red twist", "pastel spiral", "double mandelbrot", etc . . . -if no word comes to mind immediately, make anything up. -Touch fingers to the key that corresponds to the images during memorization process, i.e. touch "1" when first image is shown, etc...
I tried to assign a color and one word description to each picture, then memorized the sequence so I could recall if the picture was shown before. But when the pictures were interspaced with the letters I couldn't remember as well
Tried to remember some of the specifics of each image.
I tried to relate images with real life objects like fire, lightning, universe, earth, leaves, etc. Then try to make a story connecting them in order.
Strategies -Shutting one eye & seeing how image would look blurred. -On seeing that, try & give it a name based on how it looks -keep making a story for each block as new images are shown -chant that story while typing the letters so you don't forget -ex. "On black planet there were pieces of sea and colorful leaves, so white air god turned it into yellow lava." -Additionally, remember the 3rd image/name to keep track of numbering. -While typing letters, try to remember position for each letter shown. for example for 'k', I typed 'j' & for 'l', I typed 'k'. j-k-l are together on keyboard, so remember the position for next time

Halfway through, I tried to map the letters on the keyboard, backwards. So without having to think much I could press the correct letter. For images, mostly looked at color, symmetry and pattern.
Remember colors/associated with life events (rising to top, etc, grey -> sadness)
-remember the first 4 images & 5th one just needed to be blurred out from memory. -basically only f,u,k was posing problem initially but then I knew it appeared more than any other letter, so got hold of it.
I tried create stories when “please wait” set was on. and when the other type of block was on I tried to remember 1st and last images.
-Tried to describe the image in a couple of words and remember in that sequence. -Tried to recall shapes too when possible.
-Called each artwork a particular name -Recognized common Math patterns (elliptic curves, Mandelbrot sets, fractals, etc) -Tuned out internal abcdef. . . song for letters.