The impact of effort on the search for information

by

Karissa Payne

B.A., University of Findlay, 2019
M.S., Kansas State University, 2023

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Abstract

In everyday environments, people encounter vast amounts of information but are limited in what they can realistically use in making decisions. Because acquiring information often requires effort, individuals may favor less demanding sources even at the expense of their accuracy. However, it remains unclear how effort costs and the predictive validity of information jointly shape information search in decision-making. This study investigates how these factors interact in probabilistic environments. Across two experiments, participants explored a two-dimensional digital environment, using their mouse cursor to uncover cue information used to predict answers in a two-alternative forced choice task. Cue predictive validities ranged from 50% to 100%, and effort was manipulated through cursor speed and cue distance. Experiment 1 tested how effort costs and predictive validity influenced sampling when validities were explicitly presented on screen, whereas Experiment 2 required participants to learn predictive validities through trial and error. In both experiments, higher effort reduced sampling of less valid cues, while perfectly predictive (100%) cues were consistently prioritized even under greater effort, demonstrating a certainty effect in information search. When predictive validities were hidden in Experiment 2, effort exerted a stronger deterrent effect than it did when predictive validities were known (Experiment 1), and individual variability in exploration strategies increased. Analyses further indicated that participants treated perfectly predictive cues as categorically distinct, and the functional form of effort's influence was similar to that of a delay discounting curve. Together, these findings support a cost-benefit view of information search, in which people weigh expected information gain against the cost of effort. Effort constrains information sampling, but perfectly predictive information resists the cost of effort much more than less predictive information.

The impact of effort on the search for information

by

Karissa Payne

B.A., University of Findlay, 2019
M.S., Kansas State University, 2023

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Approved by:

Major Professor
Dr. Michael Young

Copyright

Abstract

In everyday environments, people encounter vast amounts of information but are limited in what they can realistically use in making decisions. Because acquiring information often requires effort, individuals may favor less demanding sources even at the expense of their accuracy. However, it remains unclear how effort costs and the predictive validity of information jointly shape information search in decision-making. This study investigates how these factors interact in probabilistic environments. Across two experiments, participants explored a two-dimensional digital environment, using their mouse cursor to uncover cue information used to predict answers in a two-alternative forced choice task. Cue predictive validities ranged from 50% to 100%, and effort was manipulated through cursor speed and cue distance. Experiment 1 tested how effort costs and predictive validity influenced sampling when validities were explicitly presented on screen, whereas Experiment 2 required participants to learn predictive validities through trial and error. In both experiments, higher effort reduced sampling of less valid cues, while perfectly predictive (100%) cues were consistently prioritized even under greater effort, demonstrating a certainty effect in information search. When predictive validities were hidden in Experiment 2, effort exerted a stronger deterrent effect than it did when predictive validities were known (Experiment 1), and individual variability in exploration strategies increased. Analyses further indicated that participants treated perfectly predictive cues as categorically distinct, and the functional form of effort's influence was similar to that of a delay discounting curve. Together, these findings support a cost-benefit view of information search, in which people weigh expected information gain against the cost of effort. Effort constrains information sampling, but perfectly predictive information resists the cost of effort much more than less predictive information.

Table of Contents

List of Figures viii
List of Tables x
Acknowledgements xi
Chapter 1 - Introduction 1
- The Lens Model 1
- Attention and the Search for Information 3
- The Cost of Effort 6
- Effort Shaping Decisions 9
- The Impact of Effort on the Search for Information 11

Chapter 2 - Methods 18
- Paradigm 18
- Participants 22
- General Task Procedure 23
- Experiment Procedure 25
  - Experiment 1 25
  - Experiment 2 28
- Data Analysis 29

Chapter 3 - Results 35
- Experiment 1 35
  - Sample Demographics 35
  - Raw Data Visualizations and Descriptive Statistics 35
  - Effort Transform Tests 43
  - Model Comparison 48
  - Model Parameter Estimates 50
  - Model Comparison with Predictive Validity as a Categorical Variable 54
  - Parameter Estimates and Conditional Effects of the Best Fitting Model 56
- Experiment 2 64
  - Sample Demographics 64
  - Raw Data Visualizations and Descriptive Statistics 65

Effort Transform Tests 70
Model Comparison 73
Model Parameter Estimates 75
Model Comparison with Predictive Validity as a Categorical Variable 79
Parameter Estimates with Predictive Validity as a Categorical Variable 81
Chapter 4 - Discussion 94
Effort and Predictive Validity 94
Learning vs. Known Predictive Validity 96
The Shape of Effort 98
Limitations 100
Future Directions 101
Closing Remarks 102
References 103
Appendix A - Early Pilots 116
Appendix B - Additional Raw Data Visualizations for Experiment 1 120
Appendix C - Trial Duration Visualizations for Experiment 1 122
Appendix D - Average Speed by Mouse Setting, Experiment 1 124
Appendix E - Frequentist Models, Continuous PV, Experiment 1 126
Appendix F - Frequentist Models, Categorical PV, Experiment 1 129
Appendix G - Non-100 Cue Visit Trials, Experiment 1 132
Appendix H - Additional Raw Data Visualizations for Experiment 2 133
Appendix I - Trial Duration Visualizations for Experiment 2 135
Appendix J - Average Speed Per Mouse Setting, Experiment 2 138
Appendix K - Frequentist Models, Continuous PV, Experiment 2 140
Appendix L - Non-100 Cue Visit Trials, Experiment 2 143
Appendix M - Conditional Effects for Experiment 2, by Block 144

List of Figures

Figure 1 ... 2
Figure 2 ... 13
Figure 3 ... 20
Figure 4 ... 26
Figure 5 ... 36
Figure 6 ... 37
Figure 7 ... 39
Figure 8 ... 40
Figure 9 ... 42
Figure 10 ... 44
Figure 11 ... 45
Figure 12 ... 52
Figure 13 ... 53
Figure 14 ... 59
Figure 15 ... 60
Figure 16 ... 63
Figure 17 ... 64
Figure 18 ... 65
Figure 19 ... 66
Figure 20 ... 67
Figure 21 ... 68
Figure 22 ... 70
Figure 23 ... 77
Figure 24 ... 78
Figure 25 ... 83
Figure 26 ... 84
Figure 27 ... 87
Figure 28 ... 88
Figure 29 ... 90

Figure 30 92
Figure 31 93
Figure 32 117
Figure 33 118
Figure 34 119
Figure 35 120
Figure 36 121
Figure 37 122
Figure 38 123
Figure 39 124
Figure 40 125
Figure 41 128
Figure 42 131
Figure 43 132
Figure 44 133
Figure 45 134
Figure 46 135
Figure 47 136
Figure 48 138
Figure 49 139
Figure 50 142
Figure 51 143
Figure 52 144
Figure 53 145

List of Tables

Table 1 27
Table 2 47
Table 3 49
Table 4 51
Table 5 56
Table 6 58
Table 7 72
Table 8 74
Table 9 76
Table 10 80
Table 11 82
Table 12 123
Table 13 126
Table 14 127
Table 15 129
Table 16 130
Table 17 137
Table 18 140
Table 19 141

Acknowledgements

How lucky I am to be in a position where there are too many people to thank! I will not be able to list everyone by name but know that if we've interacted at all within my tenure at Kansas State, a little or a lot, you've helped shape me into the person I am today.

I would like to thank my dissertation committee of Dr. Mike Young, Dr. Heather Bailey, Dr. Gary Brase, and Dr. Rachel Brown. Mike, thank you for shaping me into both a competent and confident professional. Even beyond the incredible wealth of stats and research knowledge you've imparted over my years in graduate school, you've given me life advice that I will carry with me forever. The strength I show as a researcher and leader is directly attributable to your influence. Heather, I first worked with you early in my graduate career, and even when I was having a rough time you gave advice that helped me keep going, improve on writing, organization, and getting projects out the door! Thank you so much for fortifying my foundation. Gary, I don't know if I will ever know as much about as many things as you do but thank you for the wealth of knowledge over multiple classes. My work is much stronger, and much more well-rounded, because of your influence. And Rachel, we only met in person once but thank you for the coffee and optimism, two things the world could truly use more of. Thank you for the outlook on research outside of my bubble.

I also want to thank everyone I have worked under through CNAP: Dr. Dan Andresen, Dr. Doina Caragea, Dr. Dave Thompson, Dr. Heather Bailey, and Dr. Kim Kirkpatrick. I wouldn't have half as many skills as I have today without the ample opportunities you provided for me to create and lead. Thank you all for taking a chance on me. Related to CNAP, thank you Beocat! Data analysis for this project would not have been remotely feasible without the Beocat

Research Cluster at Kansas State University (funded in part by NSF grants CNS-1006860, EPS-1006860, EPS-0919443, ACI-1440548, CHE-1726332, and NIH P20GM113109).

I want to thank my brain trust of cats: Mia, Mo, and Malley. Your feedback has been immeasurable and invaluable, particularly because it was all meows, chirps, and squeaks.

And finally, I want to thank my husband, Dr. Brian Howatt. Brian, thank you for all of your support over the years. Thank you for listening to me and believing in me, even when I didn't believe in myself. Thank you for making me laugh. Thank you for being a constant warm, positive presence in my life. Thank you for being a sounding board for any and all of my ideas, even the weird ones. And lastly, thank you for inspiring the effort manipulation in this study, with your remarkably slow mouse cursor speed.

Chapter 1 - Introduction

There is a wealth of information available in the world, but most often we cannot consult all available information before making decisions. Information may be difficult to obtain, and the effort required may be prohibitive. These constraints can lead us to settle for easily available information, even if it may not accurately reflect the truth. You may take a coworker's word for it that it is raining outside, instead of going out to check for yourself. You may take the first search result, or an AI-generated search summary, as a sufficient resource on a topic instead of going to the library and checking out a book from an expert. A timely subject given the advent of large language models, the impact of effort on information search has not been thoroughly researched. Within this dissertation, I aim to construct a paradigm in which effort and the validity of information can be manipulated on continuous scales, allowing us to inspect the relative impact of both factors on our search for and utilization of information. I will start with a discussion of Brunswik's Lens Model, contextualizing information search in a probabilistic world. Following this context, I will discuss attention, the mechanism by which information enters center stage in the brain, and effort, a physical and cognitive cost. This background will inform the paradigm with which I will investigate effort's impact on the information we gather in the world.

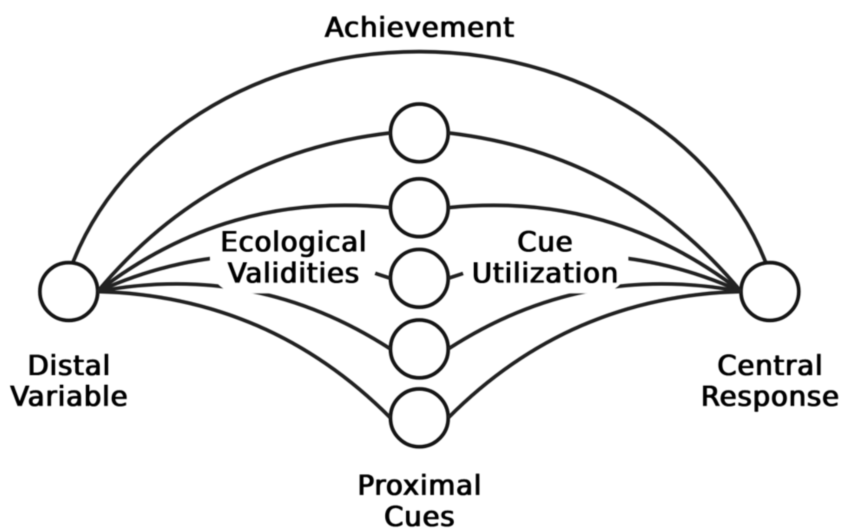
The Lens Model

Information we can gather from our environment can be considered "cues" reflective of the true state of the world. The degree to which a cue reflects reliable information is that cue's predictive validity (Cronbach & Meehl, 1955; Little & Lewandowsky, 2012). The relationship between cues, the environment, and the decision maker is captured by Egon Brunswik's Lens Model (see Figure 1). This model stems from Brunswik's general outlook of Probabilistic

Functionalism, a branch of psychological study focusing on the goals and achievements of organisms within the environments they sample information from (Brunswik, 1955; Doherty & Kurz, 1996). Organisms cannot tap directly into the truth of the world, rather they sample probabilistic information, or information that has a probability or chance of reflecting the truth of the world. An example would be a doctor observing symptoms of an illness like coughing or sneezing—the illness itself is not measurable, but the symptoms are. Influenced by colleague Edward Tolman's purposive behaviorism (Tolman, 1932), Brunswik crafted his Lens Model to illustrate achievement in a probabilistic world.

Figure 1

Brunswik's Lens Model



Note. Illustration adapted from Brunswik, 1955.

Cues, in the center of the Lens Model, are considered *proximal*—they are the information that is within grasp of the organism. The truth of the world is *distal*, on the left-hand side. The organism's actions are considered the *central* response. Proximal cues are an organism's only means of sampling the environment. The degree to which an organism samples and uses a cue in making a decision is cue *utilization*. This utilization connection is our current area of inquiry. In order to utilize a cue, one must first obtain its information. How does effort shape this process?

The degree to which information cues match the distal state of the world is called *ecological validity*. Ecological validity differs from predictive validity, defined above, in that predictive validity instead focuses on how well a cue predicts any outcome—a cue that usually predicts an incorrect outcome still has high predictive validity (i.e., you can utilize that cue to know what *not* to do). A cue that does not reliably predict any outcome has little predictive validity. Within this study, I focus on predictive validity as the measure of how useful information is to the participant. Finally, achievement (or functional validity), is the resulting degree to which the organism's response achieves the desired outcome—the ideal outcome if the organism had direct access to all possible valid information (Brunswik, 1955; Doherty & Kurz, 1996).

It is optimal for organisms to utilize cues that have the highest predictive validity—assuming the best decisions can be made with the most reliable information (Doherty & Kurz, 1996) and that all of the cues are available at the same cost. However, obtaining and using cue information is in itself a detailed process. I will start by defining how we get information from cues—through attention.

Attention and the Search for Information

Attended information is information that is presently being processed by the brain, the most active part of working memory, and what is at the forefront of our mind (Cowan, 1999, 2017; Oberauer, 2002, 2009, 2019). Attention is our cognitive engagement with information, and it is frequently matched by corresponding sensory activation, like movements of the eyes to focus on visual information (Just & Carpenter, 1976; Yarbus, 1967). As an example, you have just moved your eyes to put this sentence in the center of your vision, in order to read and comprehend the words and integrate them with the other concepts from this paragraph. The motor movement of the eyes in this case is a prerequisite for the cognitive processing of this visual information.

Attention shapes what we can learn about information. Mackintosh, in his theory of attention, referred to an item's associability as its capacity to be learned about. The more attention given to a cue, the higher its associability, that is, the more information can be learned about it (Mackintosh, 1975). In cue utilization, attention is the means by which information from proximal cues gets to the brain, shaping what information we can learn about these cues.

Historically, attention has played a limited role in decision making research, but the consideration of attention in this domain has grown in recent years (Mormann & Russo, 2021; Orquin et al., 2021). The importance of attention in shaping cue utilization has become more apparent, with an increasing number of models of decision-making behavior incorporating attentional aspects, like the Attentional Drift Diffusion Model (Krajbich et al., 2010; Mormann & Russo, 2021). An important consideration to these models and our present inquiry is: what draws attention?

Attention can be drawn by bottom-up, physical properties of stimuli. In the example of visual attention, these properties can be features like color, contrast, or motion. Features that stand out among others are said to be "salient" and can capture our attention without us consciously seeking them out (Henderson, 2007; Itti et al., 1998; Orquin et al., 2021). Attention to options in decision making can also be driven by the position and size of the options, along with the number of options there are to choose from (Orquin et al., 2021).

Attention can also be consciously directed through *top-down* processes (Baluch & Itti, 2011; Henderson, 2007). Cognitive factors like goals and task demands can drive us to attend to relevant information. In the case of visual attention, this top-down influence on attention is reflected in measurable eye movements whose locations correlate with the cognitive processing at hand (Just & Carpenter, 1976; Yarbus, 1967). In decision-making research, top-down attention often appears in the form of us looking more at the option we ultimately choose (Mormann & Russo, 2021; Orquin et al., 2021). This increased attention to chosen outcomes has spurred debate as to whether or not attention itself increases the value of a choice. Presently, Mormann and colleagues have found insufficient evidence to support this point, but attention and perceived choice value are indeed correlated (Mormann & Russo, 2021).

What is clear is that attention is shaped to focus on valuable information. In the case of proximal cues, a focus on valuable information means attention is drawn to cues that are reliably predictive. Predictive information is valuable for making adaptive decisions (Dall et al., 2005), and as such, we attend more to cues that we have learned to have high predictive validity (Le Pelley et al., 2011; Mackintosh, 1975). Increased attention to predictive cues speaks to the important implications of information search. Cues that are attended to will gain associability, and cues that prove to be predictive draw attention. Furthermore, this shift in attention comes at the cost of attention to other cues. In the case of blocking, a novel cue presented alongside a cue with learned predictive validity will be slow to gain associative strength, even though it is also predictive. The focus is already on the original predictive cue, and attention to new information from the second cue is suppressed (Kruschke, 2003; Mackintosh, 1975).

The background of attention research speaks to its importance in the everyday functioning of organisms. Attention can shape what we learn about the world, how adaptively we behave, and how we seek out future information. However, we are restricted by practical limitations when sampling the environment. We cannot perfectly sample all information available to us, as information gathering can come with costs like required effort.

The Cost of Effort

Attention itself can be effortful. Kahneman (1973) proposed that attention is equal to effort, but this claim has since been debated. Bruya and colleagues (Bruya, 2010; Bruya & Tang, 2018) have discussed how attention can happen under sympathetic dominance, where it feels effortful, and parasympathetic dominance, where it feels effortless. The distinction between effortful and effortless attention echoes the distinction between mandatory and volitional top-down attention from Baluch and Itti (2011), with mandatory attention being automatic and driven by prior knowledge (think automatically looking at the face of someone talking to you), and volitional attention entailing conscious direction of attention towards something task or goal related (think parsing faces in a crowd to find someone you are scheduled to meet with).

Even attention that is not effortful can be shaped by effort. Wickens and colleagues' SEEV model of attention suggests that visual attention is shaped by salience, what is expected, and what is valuable—but attention is also shaped by effort (Wickens, 2014, 2015). Effort in this case is broad. It can refer to eye movements themselves (eye movements are "cheap" but not "free"; Wickens, 2014), head movements (participants would prefer to memorize information than make large head movements to get the same information; Ballard et al., 1995), or other physical steps required to gather physical information (Wickens, 2014). The effort required to gather information can impact an organism's decision to do so. Effort costs and limited attentional resources means that information must be gathered in a way that maximizes value while limiting cost (Orquin et al., 2021; Sims, 2003). For these reasons, I will primarily focus on the effortful search for information, and not the more general term of attention.

Effort itself is another broad concept that requires unpacking. Effort is largely described as being unpleasant (but see Inzlicht et al., 2018 for discussion on valued effort). It can be aversive, despite its potential to lead to positive outcomes (Kool & Botvinick, 2018; Kurzban, 2016). The unpleasant feeling associated with effort is theorized to be due to motivational systems selected through evolution to guide organisms to produce adaptive behaviors, which can include not exerting excessive energy and not attending to information of little value (Kurzban, 2016; Tooby & Cosmides, 2008). However, these systems do not always drive organisms in the direction that would ultimately be most beneficial. Such motivation systems can unfortunately lead to non-adaptive behaviors as described in the self-control literature, where people and animals choose immediate gratifications (eating junk food, sitting on the couch, procrastinating to watch a movie) over long-term gains (eating healthy, exercising, completing assignments; Hagger et al., 2010; Inzlicht & Gutsell, 2007; Kurzban, 2016). Effort can be both physical (Lewis, 1964; Mitchell, 2017) as well as cognitive with increased subjective mental activity (Botvinick et al., 2009; Kool & Botvinick, 2014, 2018; Mitchell, 2017). Even when cognitive effort is beneficial, like higher germane cognitive load leading to enhanced learning (an example being learning more through taking tests compared to reading), it is still characterized by an unpleasant feeling (Paas et al., 2010; Roediger & Karpicke, 2006; Wickens, 2014).

Alongside unpleasant subjective experience, effort has been theorized to be costly due to potential energy costs. Kahneman suggested that effort takes energy from a limited store, like metabolic energy. Cognitive effort and attention would therefore be a type of metabolic expenditure in the brain (Bruya & Tang, 2018; Kahneman, 1973). Cognitive effort in particular could also take from some capacity for calculation and information processing, like working memory capacity (Anderson, 2013; Baddeley, 2007; Engle, 2002; Kool & Botvinick, 2018; Shenhav et al., 2017). Energy cost and mental resource hypotheses have been heavily debated, with a hearty amount of research arguing against both (Beedie & Lane, 2012; Carter et al., 2015; Kool & Botvinick, 2018; Kurzban, 2010, 2016; Orquin & Kurzban, 2016). Energy costs are easier to identify for physical effort and exertion (Summerside & Ahmed, 2021), but cognitive effort being tied to the utilization of a limited resource is still up for debate.

Kurzban suggests that the feeling of something being effortful might instead have more to do with the choices we make than an objective resource depletion. Effort is most often confounded with time (Białaszek et al., 2017; Kurniawan et al., 2010). Because effortful tasks take time, they come with an opportunity cost—when you do one task, you are missing out on doing other tasks that could potentially be more rewarding (Kurzban, 2016; Kurzban et al., 2013), especially if those tasks are not ones that can be completed in tandem (Wickens, 2014). When effortful tasks are pitted against less effortful alternatives with more immediate rewards, the opportunity cost could therefore lead to an aversive feeling towards continued effort exertion. Researchers like Kurzban state that opportunity costs are the primary cost of effort.

The choice to make an effort is similar to the choice between exploration and exploitation—seeking out new options to maximize future benefits, versus deriving maximum benefits from known sources (Cohen et al., 2007; Hills et al., 2015; Mehlhorn et al., 2015). Exploration and exploitation appear in the animal behavior literature through choices to exploit the present environment for resources, or to physically explore other regions for potential gain (Bautista et al., 2001; Kurniawan et al., 2010; MacArthur & Pianka, 1966). The choice of staying where one is versus venturing to other options also applies to information search and attention. Attention can be centered on one cue, alternative, or topic, or attention can be broadened to other alternatives (Bruya & Tang, 2018; Kool & Botvinick, 2018; Wilson et al., 2014). Similar to the discussion of opportunity costs, Kool et al. suggest that the cost of effort might come from an evolutionary bias that rewards exploration (Inzlicht & Schmeichel, 2016; Kool & Botvinick, 2018; Schouppe et al., 2014).

Effort Shaping Decisions

Regardless of the origin, effort shapes the decisions we make (Białaszek et al., 2017; Kool & Botvinick, 2018; Kurniawan et al., 2010). The cost of effort is factored into a cost/benefit analysis for taking action (Kool & Botvinick, 2018). We generally like to minimize the effort we exert. The Law of Less Work from behaviorism suggests that, if two behaviors are equally reinforced but differ in their effort requirement, people and animals will choose the less effortful option (Hull, 1943; Mitchell, 2017). Behavioral economics has found that if the effort cost is too high, people will choose leisure instead (Kool & Botvinick, 2014; Kurzban, 2016).

One way effort shapes decisions is through our usage of heuristics. Heuristics are mental shortcuts that reduce the cognitive effort in decision-making (Shah & Oppenheimer, 2008; Tversky & Kahneman, 1974). Shah and Oppenheimer define an effort-reduction framework, describing how these mental shortcuts reduce or eliminate steps needed to make a decision (Shah & Oppenheimer, 2008). An example effortful decision-making algorithm is the weighted additive rule where all alternatives and/or cues are considered, weighted, and then evaluated for their value (J. W. Payne, 1993; Shah & Oppenheimer, 2008). Shah and Oppenheimer define how heuristics make this process easier. Organisms can, for example, examine fewer cues, integrate less information, and simplify how the cues are weighted. Tversky and Kahneman identify three common heuristics used in assessing the probability of an outcome. People may estimate probability based on how close an event is to a typical case or stereotype (representativeness heuristic), based on how easily examples come to mind (availability heuristic), or based in initial information that skews judgements (anchoring heuristic; Tversky & Kahneman, 1974). These shortcuts are frequently, and unconsciously, implemented in order to save on the effort costs of making a multitude of decisions in our daily lives.

Decisions made in the face of costs are studied through discounting. Particularly, rewards can be devalued, or discounted, by having to wait for them (delay or temporal discounting; Green & Myerson, 2004; G. F. Loewenstein, 1988; Rachlin et al., 1991), having to work for them (effort discounting; Białaszek et al., 2017; Mitchell, 1999, 2017; Sugiwaka & Okouchi, 2004) or having them be uncertain (probability discounting; Green & Myerson, 2004; Rachlin et al., 1991). Delay and probability discounting have been heavily researched in psychology, behavioral economics, neuroeconomics, and biology. However, research in effort discounting is growing (Białaszek et al., 2017; Mitchell, 2017). Discounting is formally studied by presenting participants with options that vary in magnitude as well as the variable of interest (delay, effort, or probability). Through this method, researchers can identify the point in which a smaller magnitude option is considered indifferent from a larger magnitude option plus the additional cost of the measured variable. In the example of delay discounting, a smaller reward may be chosen if it is available sooner, versus a larger reward that is available after a notable period of time.

For effort discounting, a participant may forgo a larger reward if a smaller reward can be obtained without as much physical or cognitive effort (Kool & Botvinick, 2018; Kurniawan et al., 2010; Mitchell, 2017). Effort used in human effort discounting studies has largely been hypothetical (Mitchell, 2017). The most frequent implementation of experienced effort in effort discounting studies is in the form of a sample of effort only at the beginning of an experiment. Participants may experience an example of an effortful task, and then they may be asked to make hypothetical choices based on their memory of how effortful the example was. Such samples of effort have included physical effort of squeezing hand grips (Białaszek et al., 2017; Mitchell, 1999) and cognitive effort of task switching and N-back tasks with different values of N (Botvinick et al., 2009; Westbrook & Braver, 2015). Importantly, however, hypothetical or partially hypothetical effort discounting tasks may have results that differ from tasks where participants fully and repeatedly experience the effort of each choice (Mitchell, 2017).

The Impact of Effort on the Search for Information

How might effort discount the value of information cues? It is reasonable to anticipate information to be susceptible to the same discounting effects that can reduce the value of other choice targets. However, the discounting of information may be reduced by our strong desire for certainty, an effect that has been found in probability discounting studies and may also apply to the search for valid information

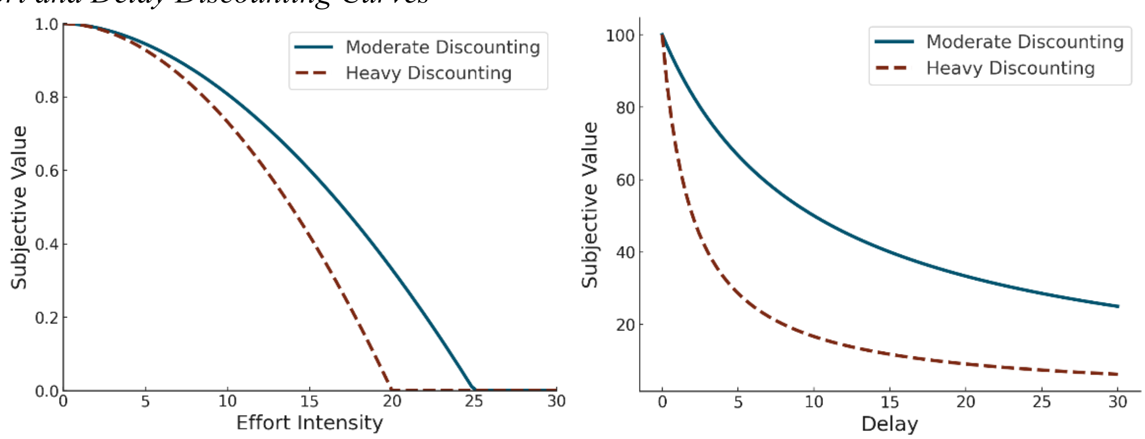
When choosing between options, people will overweight options that are certain and underweight those that are not (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Certainty can even take priority over reward magnitude. Work by Young and colleagues showed that participants were dramatically more likely to wait for rewards that had higher probabilities of occurring than those of higher magnitude (Webb & Young, 2015) and participants were more likely to wait for certainty of a reward to increase, even when the magnitude of that reward decreased (Young et al., 2014). Similar to these effects found in probability discounting studies, it is possible that information with higher predictive validity (a higher probability of reflecting the accurate state of the environment, shaping decisions leading to reward) will be overweighted, even if it is more effortful to obtain.

A certainty effect in the search for information could result in participants using a maximizing strategy, where participants always go to the most valid cue of the available options (Little & Lewandowsky, 2012; Simon, 1955, 1956; White & Koehler, 2007). Another possible strategy is probability matching, where participants respond by choosing options proportionally to their payoff probabilities. An option with 70% probability (or in the case of information search, predictive validity) may be chosen 70% of the time (Herrnstein, 1961; Little & Lewandowsky, 2012). Probability matching is the more common behavior in human decision making, even when not the most optimal strategy. However, probability matching is not as common of a strategy as maximizing in studies where participants choose between predictive cues (White & Koehler, 2007), indicating a potential larger desire for more valid information. A maximizing strategy could be exacerbated by our attention being drawn to predictive cues, to the detriment of non-predictive cues (Le Pelley et al., 2011; Mackintosh, 1975). Cue utilization can, however, be diminished through manipulations like time pressure (Rothstein, 1986), leading to the implementation of heuristics like relying on fewer cues.

The impact of an effort manipulation may differ from effects seen in delay and probability discounting studies, as the shape of the effort discounting curve has been noted to differ. There has been debate on the function that best captures the shape of the effort discounting curve, but most studies have found it to be concave (for both physical and cognitive effort), compared to the convex curves for both delay and probability discounting (see Figure 2 for comparison; Białaszek et al., 2017; Klein-Flügge et al., 2015; Mitchell, 2017). This shape is important to consider, as it defines when increasing effort begins to devalue options. In the convex shapes of delay and probability discounting (most frequently modeled with a hyperbolic function), rewards are devalued at the highest rate early on, when delay is first introduced or when probability first begins to drop. In contrast, the concave shapes seen in effort discounting curves (fit to a variety of models with no conclusive best option, including power [Białaszek et al., 2017; Hartmann et al., 2013], exponential [Białaszek et al., 2017], and hyperbolic [Białaszek et al., 2017; Sugiwaka & Okouchi, 2004]) show the highest rate of devaluation occurring towards the end of the measured range of effort. If more effort is already being expended, additional effort appears to have a greater impact (Białaszek et al., 2017; Kool & Botvinick, 2018). Klein-Flugge et al. suggest this shape to be due to the low energy cost of low effort, with effort only becoming a burden when it has passed a threshold at which it is too much to bear (Białaszek et al., 2017; Klein-Flügge et al., 2015).

Figure 2

Effort and Delay Discounting Curves



Note. Left: example effort discounting curves, adapted from Białaszek et al., 2017. Right: example delay discounting curves, adapted from Gray & MacKillop, 2015.

In this dissertation, I sought to elucidate the effect effort can have in shaping information search. Information has high value, with broad implications for what we learn (Just & Carpenter, 1976; Yarbus, 1967) and the information we seek out in the future (Le Pelley et al., 2011; Mackintosh, 1975). How the human mind handles hurdles in the search for information requires further study. I had three sets of hypotheses for this study. The first set (Hypotheses 1A, the Effort Dominance Hypothesis, and 1B, the Predictive Validity Dominance Hypothesis) addresses which competing factors may or may not have an effect on the search for information.

Hypothesis 1A, the Effort Dominance Hypothesis: An increase in effort may discount the value of information cues, leading participants to seek out less effortful cues even when they have a lower predictive validity than more effortful cues. This hypothesis is suggested by the cost of effort that has shown to shape behavior in humans (Białaszek et al., 2017; Hull, 1943; Kool & Botvinick, 2014, 2018; Kurniawan et al., 2010), including shaping the decisions made in effort discounting studies (Białaszek et al., 2017; Mitchell, 2017).

Hypothesis 1B, the Predictive Validity Dominance Hypothesis: It is also possible that our need for certainty drives us to seek out the information most predictive of the true state of the world, even at higher effort costs. This hypothesis would be suggested by high value of predictive information in making decisions (Dall et al., 2005), our attention system being shaped to focus on predictive cues (Le Pelley et al., 2011; Mackintosh, 1975), the certainty effect seen in probability discounting studies (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992; Webb & Young, 2015; Young et al., 2014), and the maximizing strategy (opting for the best available choice) found to be used more frequently by humans in studies where subjects choose between predictive cues (White & Koehler, 2007).

Importantly, Hypotheses 1A and 1B are not entirely mutually exclusive. It is possible that both of these effects may impact the search for information as the cost of effort and cue predictive validities change in our environment. This study will examine the degree to which both factors of effort and predictive validity shape information search, and how the effect of these factors may change in different scenarios.

The second set of hypotheses (2A, the Effort Shaping Cue Learning Hypothesis, and 2B, the Learning Similarly Despite Effort Hypothesis) apply to a particular method by which effort may shape the search for information—by altering which information is learned about and subsequently attended to.

Hypothesis 2A, the Effort Shaping Cue Learning Hypothesis: Effort will have a larger effect, and predictive validity will have a smaller effect when the cue predictive validities must be learned over time. Given that attention is drawn to cues with a high predictive validity and attention shapes what we can learn about cues (Le Pelley et al., 2011; Mackintosh, 1975), cue visits early on in the task will shape future cue visits, potentially to the detriment of cues that are not immediately identified as having a high predictive validity. In this way, effort may shape what is learned about the information landscape, resulting in a stronger impact on which cues are visited. This hypothesis is also consistent with findings from the description-experience gap in risky choice. When people learn outcome contingencies through sampling, limited and uneven exposure leads to systematic distortions in what is learned (Hertwig & Erev, 2009). Analogously, in this task, effort may constrain sampling of cues, amplifying early attentional biases and reducing the effective influence of predictive validity.

Hypothesis 2B, the Learning Similarly Despite Effort Hypothesis: It is possible that effort and predictive validity might have similar effects when predictive validities must be learned and when they are given to participants. This would be most likely to occur alongside Hypothesis 1B, the Predictive Validity Dominance Hypothesis, where the need for certainty drives behavior. If effort does not inhibit cue visits early on in learning the predictive validities, it can be anticipated that participants will learn accurate (or close-to-accurate) predictive validities for each cue and allocate their cue visits accordingly.

The third set of hypotheses (3A, the Largest Discounting Late Hypothesis, and 3B, the Largest Discounting Early Hypothesis) regards the shape of effort's discounting effect on the search for information. While the functional form of effort's discounting effect is still debated (Białaszek et al., 2017; Mitchell, 2017), it is important to investigate how varying degrees of effort may shape the value of our variable of interest in this study, and one that has not been examined with an effort manipulation before: information search.

Hypothesis 3A, the Largest Discounting Late Hypothesis: The shape of effort's effect on information search will be concave. This would match the shape found for some other effort discounting studies (Białaszek et al., 2017; Klein-Flügge et al., 2015; Mitchell, 2017), where the highest rate of devaluation occurs at higher values of effort. This hypothesis would imply that effort would not devalue information search much until effort levels are high, at which point devaluation happens more rapidly.

Hypothesis 3B, the Largest Discounting Early Hypothesis: The shape of effort's effect on information search will be convex. This would match what has been found in previous probability and delay discounting studies (Green & Myerson, 2004; Rachlin et al., 1991), where the highest rate of devaluation occurs at the lower values of the manipulation. In this study, a convex shape of the effort function would mean that the presence of effort immediately devalues the search for information, and increasing levels of effort does not increase this rate of devaluation by much.

In consideration of these three sets of hypotheses, this study used two experiments to examine the effect of effort's discounting of information, alongside comparing the relative impact of effort and predictive validity when predictive validity is known and when predictive validity must be learned.

Chapter 2 - Methods

Paradigm

Here I have constructed a paradigm that captures the probabilistic nature of the real world, while being easy for other researchers to use for their own research and replication. Representative design, championed by Egon Brunswik (Brunswik, 1955), emphasizes the importance of experiments capturing the ecological variations that shape our decisions. Here, I have implemented representative design through an experience-based decision making (EBDM) task. I varied both effort and predictive validity on continuous scales, and participants experienced real effort required to obtain the information cues. The latter is especially important, given the potential that real and hypothetical effort tasks may result in different discounting functions (Mitchell, 2017).

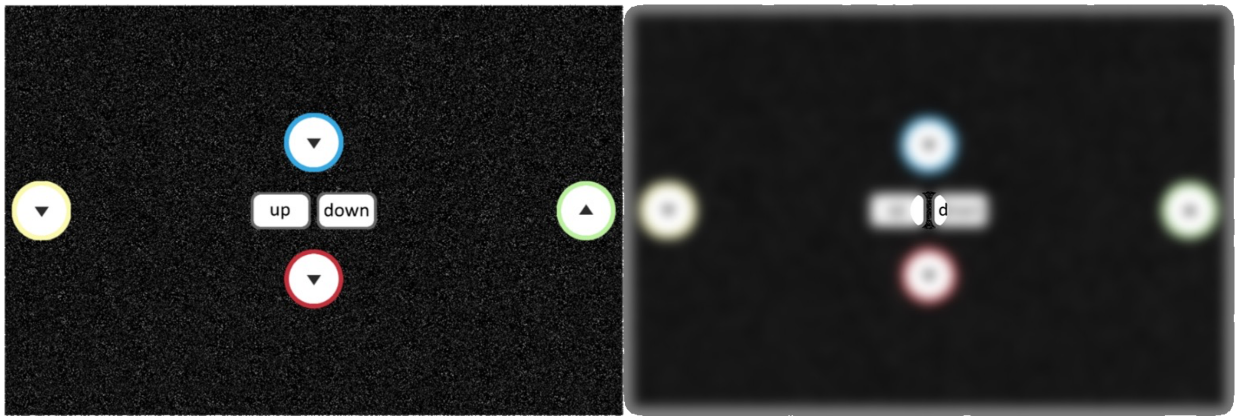
The paradigm is an alteration of a Multiple Cue Probability Learning (MCPL) task. In MCPL, cues are presented that vary in how well they predict an outcome. Participants can choose which cues and how many to use in answering what they believe an outcome will be. Through this method, researchers can study how well people learn and use probabilistic cues (Little & Lewandowsky, 2012). The primary changes to this paradigm for this study were the implementation of an effort manipulation and a measure of which cues are sought out and attended to.

Both the effort manipulation and the measure of attention were addressed through the use of a Mouse-Contingent Bi-Resolution Display. This method blurs the participant's screen, all except for a high-resolution window that is centered on the participant's computer mouse location. To unveil content on the screen, the participant needed to move their mouse to that location, providing an approximation of participant eye movements and a measure of which content has been revealed. This methodology matches that from Payne (2023) and stems from the work of Anwyl-Irvine et al. (2021), Blackwell et al. (2000), Jiang et al. (2015), Jones and Mewhort (2004), and Kim et al. (2017).

On each trial, participants were presented with four visual cues that gave information that could be used to answer a question in the center of the screen: is the correct answer "up" or "down?" Each cue circle contained an arrow pointing either up or down (see Figure 3A). With the start of each trial, the participant only saw the blurred display shown in Figure 3B. The participant needed to move the clear window over to each cue in order to unblur it and identify its information.

Figure 3

General Experiment Display Design



A. B.

Note. A. The underlying design of the experiment, with four cues providing information to aid in making the decision at the center of the screen. Participants only saw the blurred screen, shown in B, and cue information was unveiled by moving the clear mouse cursor window over to the cue location.

These cues and the correct answer varied on each trial as they would in a MCPL task (Little & Lewandowsky, 2012; White & Koehler, 2007), with each cue having a predetermined predictive validity between 50% and 100% experienced over the course of one block of trials.

The required effort to get the cue information was manipulated in two ways. A global effort variable was changed by altering the speed of the mouse cursor. The slower the mouse cursor relative to the participant's average mouse speed, the more tangibly effortful the mouse movements became. Participants could have chosen to move their mouse with more forceful, faster hand movements to cover a similar amount of ground per unit time to maintain their average movement speed, or they could have settled for the slower journey to the cue information. As with other effortful tasks, this mouse movement task is one that can coincide with delay. However, effortful tasks requiring more time to complete is a common, ecologically valid experience (Białaszek et al., 2017; Kurniawan et al., 2010). More physical exertion could get the task done faster, but a slower, less exertive task can still be perceived as "effortful," likely attributable to the opportunity cost of the task inducing a feeling of effort (Kurzban, 2016). In this way, this effort manipulation is akin to pushing a heavy box vs. a light box. A heavy box can be pushed as quickly as a light box, but a more forceful push is required.

The local effort manipulation, a variable that could be changed for each individual cue, was the cue's location relative to the center of the screen. As seen in Figure 3, some cues appeared closer to the center of the screen, while others were farther away. This distance manipulation resulted in some cues being easier to obtain while others required more work. This effect is compounded by the global effort manipulation, resulting in an overall effort term of the distance from the center of the screen (in units relative to the screen height) divided by the mouse speed coefficient.

The experiment display was presented at a size proportional to the participant's screen size, with each cue circle presented with a diameter of 11% of the screen height, and the clear window presented with a diameter 5% of the screen height (when the clear window fully entered the cue circle, the cue information in the center of the circle was visible). The blur level was set using guidance from Blackwell et al. (2000), Jones and Mewhort (2004), and Kim et al. (2017) to reach a level that obscured the task-relevant information. This level is a proportion of the participant's screen height tested to obscure the cue content on all screen sizes (a Gaussian blur with a radius of screen height / 65). This proportional design does not require the strict experimental setup used to match blur levels to the participants' visual periphery in Payne

(2023). Rather, this design was task driven. The speed of the participants' mouse movements was similarly proportional to their normal mouse speed, with the effortful condition reducing speed down to as much as 20% of its original speed. Information on screen sizes and mouse speeds was collected from the participants but was considered systemic heterogenization, or controlled variation allowed in the experimental design (von Kortzfleisch & Richter, 2024). These variables could be used in later exploratory analyses but were not the primary variables of interest in this study. Allowing for the variance in participants' laptop and desktop setups was a purposeful choice, to allow for easier implementation and replication of this paradigm by other researchers.

Participants

This study was conducted online, in order to recruit a large and diverse participant pool, alongside testing the methodology for easy online replication and further study by other labs. A power analysis was run using data from the first 30 participants from Experiment 1, using the mixedpower R package by Kumle et al. (2021). This package uses simulation techniques to perform power analyses for generalized linear mixed effects models. A-priori power analyses are a recommended method for determining the sample size required to achieve a desired error rate for effect sizes (with 0.05 as a common baseline, accepting a 5% chance of Type I error; Lakens, 2022). Simulation-based methods are recommended for complex models like the mixed effects models used in this study (Kumle et al., 2021). This power analysis suggested a sample of 30 subjects to be sufficiently powered for all tested effects and interactions in Experiment 1 (power >95% for all), using a model with random intercept effects for each participant. With the expectation that the effort effect would be stronger in Experiment 2, a target sample size of 30 was chosen for Experiment 2 as well.

General Task Procedure

The study was constructed to minimize potential confusion and disinterest from online participants. Following the recommendations from Rodd (2024), task instructions were detailed with example images from the task, alongside information about the importance of the task and the engagement expectations prior to the start of the experiment. The interactivity and game-like nature of the task was also chosen to motivate participants (Rodd describes entertainment and game-based tasks as positively influencing participant recruitment and engagement during a study; Rodd, 2024).

The size of the displayed screen was recorded throughout the experiment, to ensure the full-screen window of the study was not minimized. Four participants who left the full-screen window had their data removed from analysis (one participant from Experiment 1, and three participants from Experiment 2). I additionally checked participant activity through the recorded x-y coordinates of mouse movements, to remove trials if participants displayed excessive inactivity (Rodd, 2024). This task allowed participants to take short breaks between blocks. Inactivity greater than 2 min during an active trial was deemed as the "inactive" cutoff. No participant's data needed to be removed for this reason.

Participants were recruited from Prolific, an online participant recruitment platform with a good reputation for having engaged participants who provide high-quality data in measured attention, comprehension, and honesty (Peer et al., 2021). Participants were compensated for their time at a rate of $10 per hour. The procedure and compensation were evaluated and approved by Kansas State University's Institutional Review Board. To complete this study, participants were required to use a laptop or desktop computer.

From Prolific, recruited participants were first given a link to a Qualtrics page with the informed consent information, study description, and basic demographic questions (age and sex). Following the Qualtrics page, participants were directed to the online JavaScript experiment created with PsychoPy on Pavlovia.org (Peirce et al., 2019). Participants were instructed to not exit the full-screen experiment window during the experiment.

The Pavlovia experiment first presented participants with detailed instructions and illustrations. After this, participants moved to a practice block where they learned to visit the cues with the clear window in order to identify their predictive validities and use this information to make decisions in the center of the screen. The practice trials repeated until the participant was trained to a criterion of selecting the correct answer three times in a row.

Each trial presented cues with predetermined predictive validities (implemented with the cues being correct that percentage of the time, but the cue correct/incorrect order being randomized). The number of trials and any other unique procedure information will be described under the headers for Experiments 1 and 2. To encourage effortful responding, incorrect responses were penalized with a 2 s inter-trial interval (ITI).

After completing the task, participants were directed to a Qualtrics survey to be debriefed and to answer survey questions about their computer usage and hardware setup (recommended as per Rodd, 2024, to detect any potential issues in experiment deployment). These questions included rating their experience with computer-based video games, rated on a scale from 0 (no experience) to 5 (extensive experience). Participants were also asked if they used a laptop or desktop and were asked to identify their operating system and internet browser. Finally, participants were invited to share any issues they may have faced through the study, whether or not they were confused by any of the instructions, and whether they had feedback regarding their participation in the study. Four participants reported technical issues (three in Experiment 1, and one in Experiment 2) and were removed.

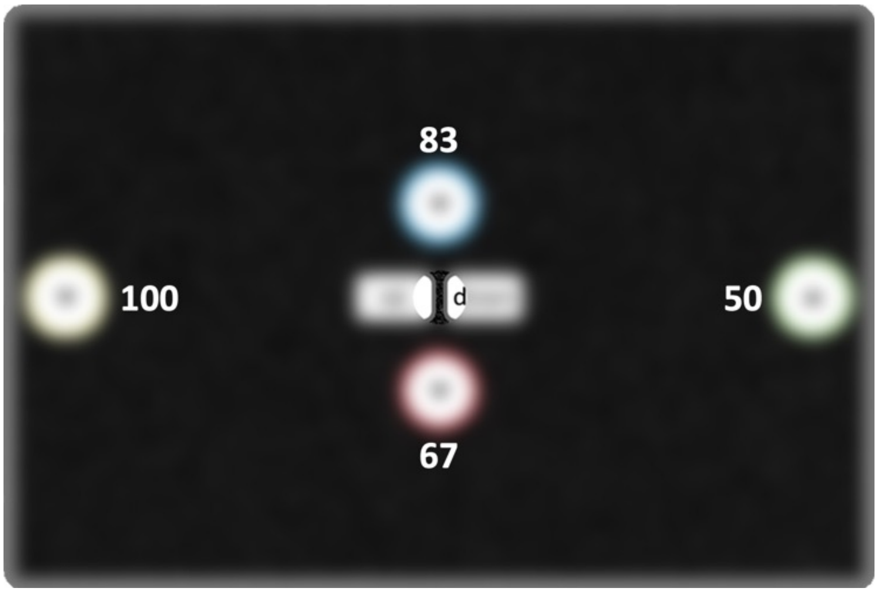
Experiment Procedure

Experiment 1

Experiment 1 presented participants with the four-cue paradigm illustrated in Figure 3, but with cue predictive validity percentages overlaid on top of the blur layer (visible without requiring mouse movements) as shown in Figure 4. Presenting the cue predictive validity percentages allowed investigation of the effect of effort on information search without the additional step of participants needing to learn the predictive validity of each cue (this learning may itself be shaped by the effort manipulation—this possibility is explored in Experiment 2). This simplification also served as a baseline measure of the strength of the effort manipulation.

Figure 4

Experiment 1 Display with Visible Cue Predictive Validities



Participants experienced 270 trials, based on an estimate of 10 s per trial, comprising 45 min of experiment trial time, leaving 15 min for instructions, practice trials, end survey, and debriefing. These time estimates and other paradigm choices were derived from early proof-of-concept pilot studies, the details of which can be viewed in Appendix A. Cue predictive validities on each trial were randomly selected from 50% (a non-informative cue), 67%, 83%, and 100% (a maximally-informative cue), such that 45 trials were allocated to each of the 6 possible combinations of the four predictive validity values assigned to near or far cues (randomizing between up/down for near cues and left/right for far cues). The global effort on each trial was randomly selected from 100% speed (the participant's standard mouse speed, little effort), 80% speed, and 20% speed (highest effort), sampling two values on the lower end of effort to provide a better spread of values of combined effort terms when we incorporate the compounding effect of the local effort manipulation (see Table 1). The local effort manipulation (cue location) stayed static throughout this experiment, with two cues close to the center (at a distance of 60% of the screen height), and two cues far from the center (at a distance of 150% of the screen height). This cue location design ensured that the "near" cues were always closer than the "far" cues, from any cue location on the screen. The resulting combined effort terms for the global and local effort manipulations are featured in Table 1. Of the 45 trials for each of the 6 cue configurations, 15 were experienced under each of three different mouse cursor speeds. The order in which these cue configurations and mouse cursor speeds were encountered was randomized for each participant. Participants were instructed that they may view as many or as few cues as they would like on each trial.

Table 1

Global and local effort settings, and the resulting combined effort terms

Mouse speed (percent of normal speed)	Cue distance from center of screen (percent of screen height)	Combined effort term (mouse speed / distance)
100	60	1.67
100	150	0.67
80	60	1.33
80	150	0.53
20	60	0.33
20	150	0.13

Note. Smaller combined effort term values correspond to higher effort

Experiment 2

Experiment 2 followed the general procedure from Experiment 1 but removed the cue predictive validity numbers from the display. Removing the visible predictive validities allowed the measurement of the relative effects of effort and predictive validity on information search when predictive validity needed to be learned through gathering cue information. As attention to cues shapes their associability, or potential to be learned about (Le Pelley et al., 2011; Mackintosh, 1975), higher effort in early trials could prevent participants from learning the value of high predictive validity cues. This experimental design is more ecologically valid, but the effects of effort and predictive validity are harder to ascertain in a one-hour experiment if participants must learn through trial and error.

Participants experienced 6 blocks of 45 trials each (anticipating up to 10 s per trial, comprising 45 min of experiment trial time, leaving 15 min for instructions, practice trials, end survey, and debriefing). Cue predictive validities on each trial were randomly selected from 50% (a non-informative cue), 67%, 83%, and 100% (a maximally informative cue), such that each block had a different combination of cue predictive validities to cue distances (near or far). The global effort in each block was randomly selected from 100% speed (the participant's standard mouse speed, little effort), 80% speed, and 20% speed (highest effort), such that each participant experienced two blocks of each of the three speeds. The predictive validity cue configurations (6 different possibilities) and the speed allocations to each block (3 different speed possibilities) were randomly assigned to each participant such that each combination was experienced at least once by one participant.

When a new block began, participants were informed with a message announcing the beginning of a new block and reminding participants that cue reliabilities may have changed.

Colored circles around the cues also changed in color to highlight the fact that the cues were new cues in a new scenario. This color change was designed to minimize confusion between cues and any bias against cues that were previously non-predictive in the preceding block (Le Pelley et al., 2011; Mackintosh, 1975).

Data Analysis

The primary goal of data analysis for this study is to estimate the relative effects that effort and predictive validity have on the search for information. To obtain these values, I used a multilevel logistic regression with effort and predictive validity predicting the outcome variable of whether or not a participant visited a particular cue on a trial (coded as a 1 for a visited cue, and a 0 for a non-visited cue). This methodology comes from Wileyto et al. (2004) and Young (2018) as an adaptation of a logistic regression method of analyzing discounting data.

Discounting data (from delay, probability, and effort discounting) are traditionally analyzed by gathering indifference points—points where participants are indifferent towards receiving a smaller immediate reward or a larger delayed reward (Białaszek et al., 2017; Kool & Botvinick, 2018; Mitchell, 2017; Wileyto et al., 2004; Young, 2018). A nonlinear regression is used to fit a function to these indifference points—commonly the hyperbolic (Rachlin et al., 1991) or hyperboloid (Green & Myerson, 2004) discounting functions—and the value of the discounting rate k is derived from this function. Higher values of k indicate a stronger preference for immediate (or more probable, or less effortful) rewards. Wileyto et al. (2004) and Young (2018) describe how this discounting value can also be obtained through the use of a logistic regression.

The logistic regressions described by Wileyto et al. (2004) and Young (2018) for delay discounting predict binary choice data from two main effects: a magnitude ratio factor and a delay ratio factor (with no intercept value β_0 in the equation, because the predicted choice with no effect of magnitude or delay should be 0.5, or chance). In Wileyto et al. (2004), the magnitude ratio factor is 1 – (magnitude of Larger Later option / magnitude of Smaller Sooner option) and the delay ratio factor is simply the delay to the Larger Later option because the delay to the smaller sooner option is 0. Young recommends transforming this factor to ln(magnitude of Larger Later / magnitude of Smaller Sooner) and the delay ratio factor to ln(delay of LL + 1) to address participants' logarithmic sensitivity to these factors as anticipated from psychophysics research (Fechner, 1948; Young, 2018). A log transform linearizes the relationship between the predictor and log-odds outcome probabilities prior to running the regression (meeting an assumption of logistic regression, see Harris, 2021 and Stoltzfus, 2011). The discounting parameter *k* can be estimated as the regression weights (or the participants' sensitivity) for the effect of delay divided by that of the effect for magnitude. However, in this study I will be evaluating the individual effects of magnitude and delay, and the credible intervals for these estimates. The ability to examine these two effects independently is important, as similar *k* values can come from different combinations of sensitivities to magnitude and delay (e.g., a participant may have a high k value because they are very sensitive to delay, or not sensitive at all to magnitude). This analysis thus addresses an existing problem in the analysis of effort discounting data—Kool and Botvinick (2018) mention the difficulty in finding the best discounting function for effort because individuals have unique capacities for effort and unique reward sensitivities (later, I will touch more on individual differences).

I use this methodology with slight adjustments to fit this paradigm. The core of the analysis is a comparison of the main effects of delay and magnitude. Here, I replace these main effects with the variables of effort and predictive validity. The effort and predictive validity of each cue on each trial are recorded, alongside the participant response of whether they visited that cue within the trial (coded as a 1) or they did not visit that cue (coded as a 0). This provides a holistic look at the effect of effort and predictive validity on cue-based decisions on a per-trial basis, without the added complexity of reevaluating information gain from individual cues and relative effort with each cue visit. Because multiple cues can be visited in a single trial, and I am not evaluating step-by-step cue visit choices but rather the end results of the trial (each cue visited or not), effort and predictive validity will not be entered as ratio terms but rather each cue's individual pre-set parameters for effort and predictive validity (the effort term being the distance of that cue from the center of the screen divided by the mouse speed coefficient—to view visualizations of the raw data separated by distance and speed, see Appendix B for Experiment 1 and K for Experiment 2). The relative effects of effort and predictive validity will be used to address Hypotheses 1A and 1B.

The predictor variables in this study are log-transformed before they are entered into the regression, as was done in Young (2018), although other transformations are explored particularly for the effort variable to see what best fits the data prior to analysis. Effort discounting curves (see Figure 2) suggest a different type of sensitivity to effort, which could influence which transformation linearizes the relationship between this predictor and the log-odds outcome variable. This possibility is explored in preparing the effort predictor variable for analysis, and the resulting functional form identified for effort addresses Hypotheses 3A and 3B. These effects are also compared between Experiments 1 and 2, in addressing Hypotheses 2A and 2B.

Differing from Wileyto et al. (2004) and Young (2018), this model has an intercept parameter, as this paradigm is new and we cannot assume that the predicted likelihood of viewing a cue is 0.5 (50/50 chance) when the predictor terms are equal to zero. Because the predictors are log transformed, a value of 1 is added to the predictors before transformation to avoid taking the log of zero (which is undefined). The variable of "trial" is also added as a main effect and interaction term, in order to evaluate the potential for the effects of effort or predictive validity to vary over the course of the experiment (further addressing Hypotheses 1A and 1B). Experiment 2 has an additional predictor of "block," to address variability in the data that could come from the trials being separated into blocks of 45 trials, over which participants learn one set of cue validities.

With this study having a within-subjects design, data gathered from one participant is not independent from other data gathered from that same participant. Multilevel modeling is used to handle this non-independence in the data (Brauer & Curtin, 2018; Matuschek et al., 2017). Not accounting for this non-independence can lead to inflated type I error rates, or false positives in results (Brauer & Curtin, 2018). In a multilevel model, the variable of "participant" is included as a random effect, as it is a variable under which other results are nested. Random effects can include intercepts and random slope effects. A random intercept for participant would allow each subjects' intercept value (when all predictors are set to zero) to differ, while a slope effect for participant would allow the slope of one of the main effects to vary across participants. The definition of the random effects structure is important. Too simple of a random effects structure can lead to inflated Type I error, but too complicated of a random effects structure can lead to a decrease in statistical power. Matuschek et al. (2017) advocate for the use of model selection strategies to identify the model best supported by the data. In this case, AIC values (Akaike, 1974) were used for model selection. I chose AIC over BIC (Schwarz, 1978) because BIC has a stronger penalty for complex models, resulting in a choice that is less likely to be an overfit model, but more likely to be an underfit model (Vrieze, 2012). In this study, each random effect tested has theoretical backing (individual differences in effort discounting [Kool & Botvinick, 2018] and sensitivity to predictive validity [White & Koehler, 2007], as well as individual differences in sensitivity to the duration of the experiment and the potential to want easier "rest" trials at different points in time [Mitchell, 2017]), so the less conservative AIC is more likely to capture real effects.

The maximal model that was investigated is as follows in R syntax (Equation 1; Note: effort may not be log transformed, as effort's functional form may be better linearized by a different transform. These possibilities were explored for model fit as well as addressing Hypotheses 3A and 3B):

$$\begin{gathered}\text{Model} <\!\!- \text{glmer}(\text{visited} \sim \log(\text{Effort} + 1) \times \log(\text{PV} + 1) \times \text{Trial} + \\ (\log(\text{Effort}+1) \times \log(\text{PV}+1) \times \text{Trial} \mid \text{Participant})\end{gathered} \tag{1}$$

Alternative models run for model comparison with AIC included those removing variations of the random slope effects, random intercept effects, and interactions in the random effects structure. In cases where the maximal model does not converge, Bayesian modeling methods were used and compared with WAIC values, an analogous method to AIC for Bayesian models (Vehtari, Gelman & Gabry, 2017).

All analyses were run in R (version 4.4.2). Frequentist multilevel models were run using the lme4 library (Bates et al., 2015) with the emmeans library (Lenth, 2019) for calculating estimated marginal means and conducting post-hoc tests. Bayesian models were run using the brms package (Burkner, 2017), and emmeans was similarly used for contrast tests on any Bayesian models. Effect sizes in all Bayesian models were determined using directional Bayes

Factors (*BF*; Makowski et al., 2019). A Bayes Factor of 20 or greater indicates strong evidence for the tested effect (Jarosz & Wiley, 2014).

Chapter 3 - Results

Experiment 1

Sample Demographics

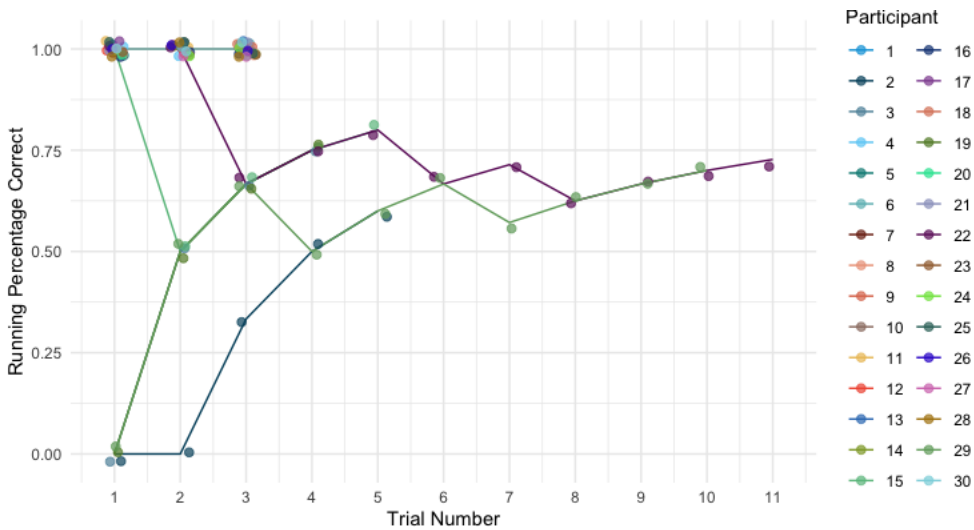
Experiment 1 originally recruited 35 participants. Of these, three were removed for reporting technical issues or issues with data saving, and one was removed for the possibility of exiting the full-screen window during the study (reported screen size decreased partway during the experiment). Additionally, one participant was removed who reported attempts to circumvent the blur. Subjects were recruited until a final sample of 30 was reached. The sample for Experiment 1 had 15 female and 15 male participants, with a mean age of 40 (*SD* = 9).

Raw Data Visualizations and Descriptive Statistics

Participants in Experiment 1, both in practice trials and the full experiment, primarily visited and relied only on the 100% cue (a strong strategy, and one that is not inhibited by having to learn cue predictive validities through trial and error, as in Experiment 2). Most participants completed the practice round requiring three correct answers in a row within three trials (26/30 participants; see Figure 5), and all participants finished the practice within 11 trials. No participant got through the practice round by solely guessing and visiting no cues (see number of cue visits per trial for the practice round in Figure 6).

Figure 5

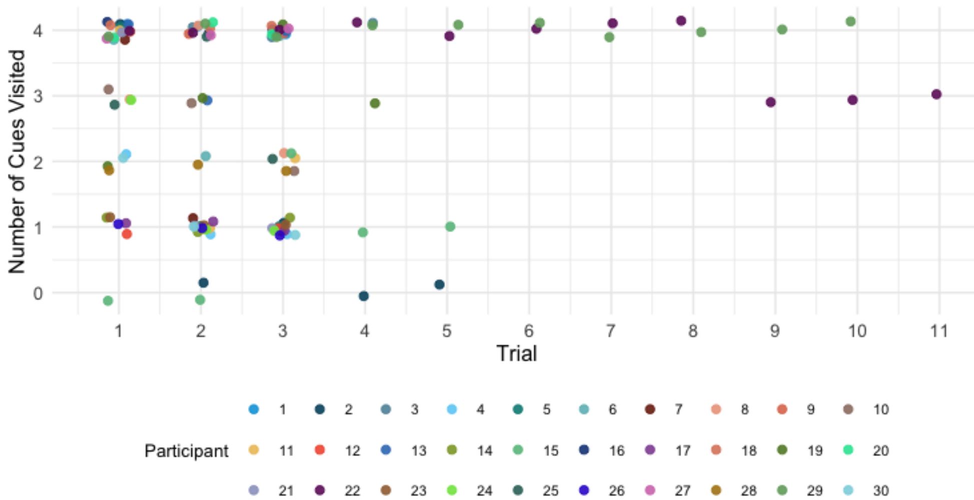
Running Proportion of Correct Responses from Participants in Practice Trials of Experiment 1



Note. The line representing each participant's running proportion of correct responses will deflect downwards when that participant makes an incorrect response.

Figure 6

Number of Cues Visited by Participants in Practice Trials of Experiment 1

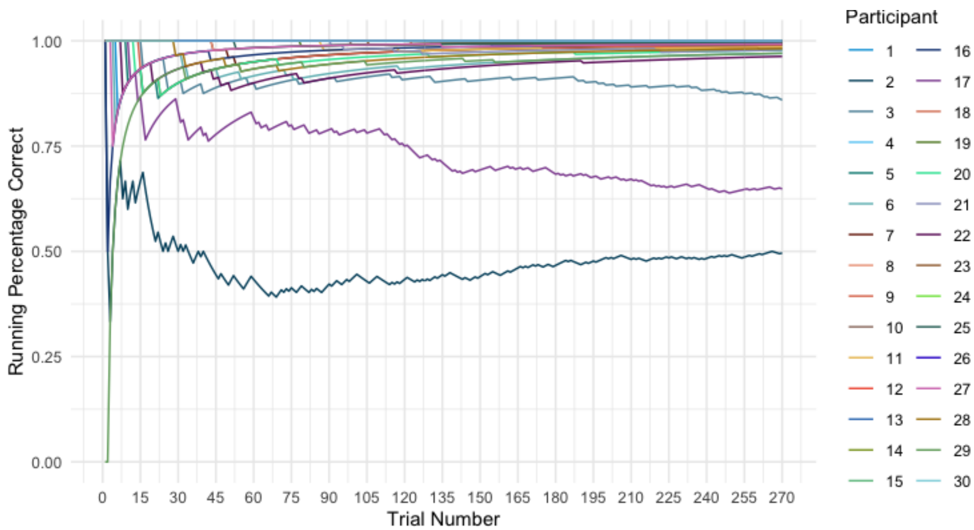


Note. Number of cues visited by each participant in the practice trials for Experiment 1. Practice trials continue until the participant gets three responses correct in a row.

In the full study, most participants were able to maintain a running percentage of correct responses of 86% or higher by Trial 15 (28 of 30 participants, see Figure 7). Only two participants seemed to have exhibited a strategy of guessing without cue visits, resulting in a running percentage of correct responses that centered around or drifted towards 50% (chance; see Figure 7). These two participants are also visible as frequently visiting no cues in a visualization of the number of cue visits per trial, Figure 8. Figure 8 also shows that the large majority of subjects began to rely on only one cue as the experiment progressed. Visiting only one cue on average led to the shortest duration of trial plus potential penalty ITI, (*Mdn* = 6.06 s, *IQR* = [4.95, 7.83]), suggesting that participants may have adopted a one-cue-visit strategy (and visiting only the 100% cue) to minimize experiment duration. Conversely, visiting no cues and just guessing between the two answers resulted in the highest trial plus ITI duration (*Mdn* = 11.55 s, *IQR* = [3.49, 17.97]), likely as guessing would result in a 2 second penalty ITI ~50% of the time. The average duration of each trial plus ITI is plotted in Appendix C, by the number of cues visited on each trial.

Figure 7

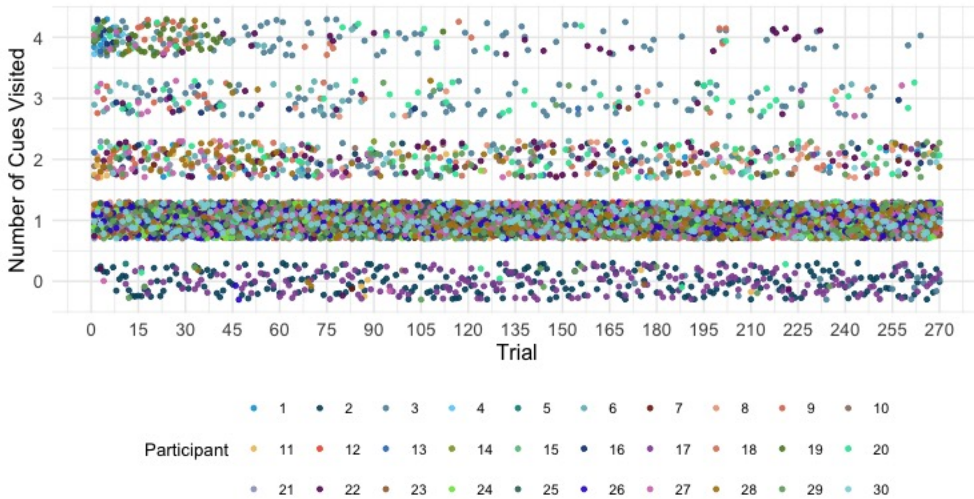
Running Proportion of Correct Responses from Participants in Experiment 1



Note. The line representing each participant's running proportion of correct responses will deflect downwards when that participant makes an incorrect response.

Figure 8

Number of Cues Visited by Participants in Experiment 1



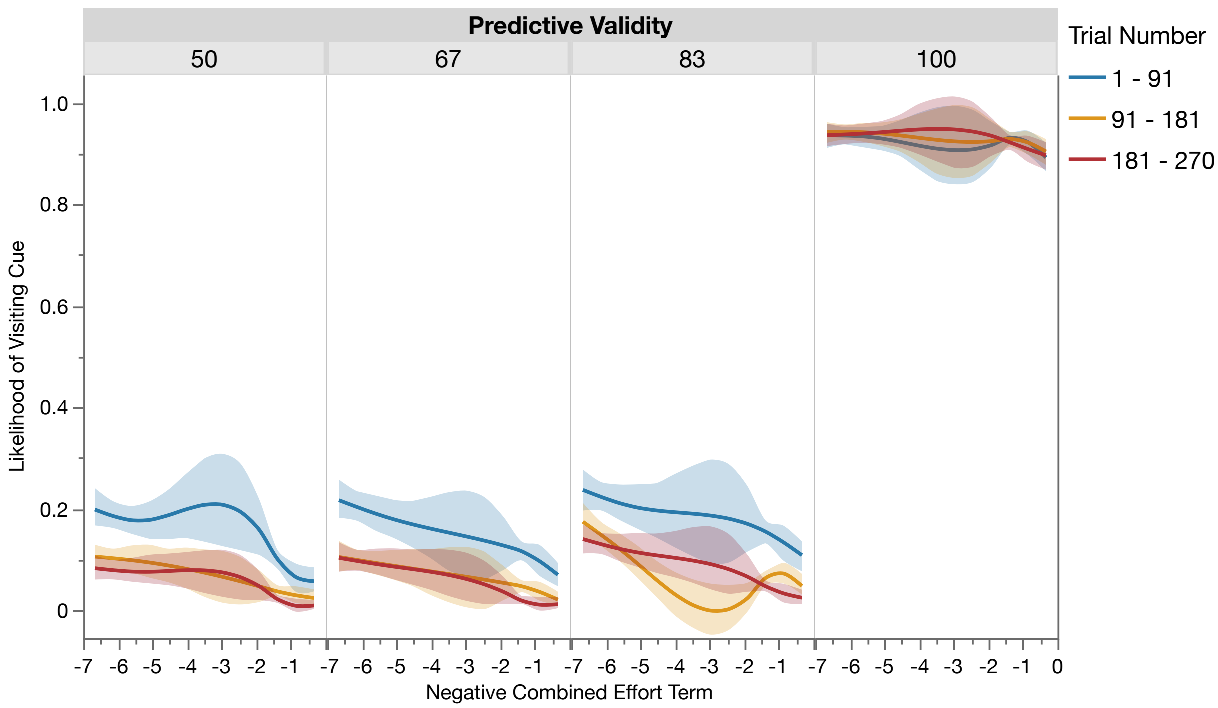
Note. Number of cues visited by each participant in each trial of Experiment 1. Over the course of the experiment, participants gravitate towards viewing only one cue.

A general desire to finish the experiment faster was also evidenced by participants moving their mouse cursors with greater physical movements when the mouse speeds were manipulated to be slower. Visualization of the outcome mouse movement speeds under the different speed coefficient settings can be seen in Appendix D. When the mouse speed was set to 80% of its original speed, participants moved at a median speed of 0.41 screen height units/s (*IQR* = [0.25 , 0.59]). This was 87% of their median for 100% speed (*Mdn* = 0.47 screen height/s, *IQR* = [0.30, 0.68]), achievable by moving the mouse or trackpad physically more in a shorter amount of time than was done at the 100% speed setting. Participants appeared to make even larger physical movements to compensate for the slowest mouse speed of 20%, with participants moving a median of 0.16 screen height/s (*IQR* = [0.09 , 0.25]), which was 34% of the median for 100% speed.

The raw data for cue visits is visualized in Figure 9, separated by the predictor variables of interest. This raw data immediately shows a distinct difference in responses for the 100% cue compared to all others. This is an important factor that will shape the analyses. Additional visualizations of this raw data, including data separated by Participant and data separated by the individual effort factors of mouse speed and distance, can be seen in Appendix B.

Figure 9

Raw Data for Experiment 1 by Effort, Predictive Validity, and Trial



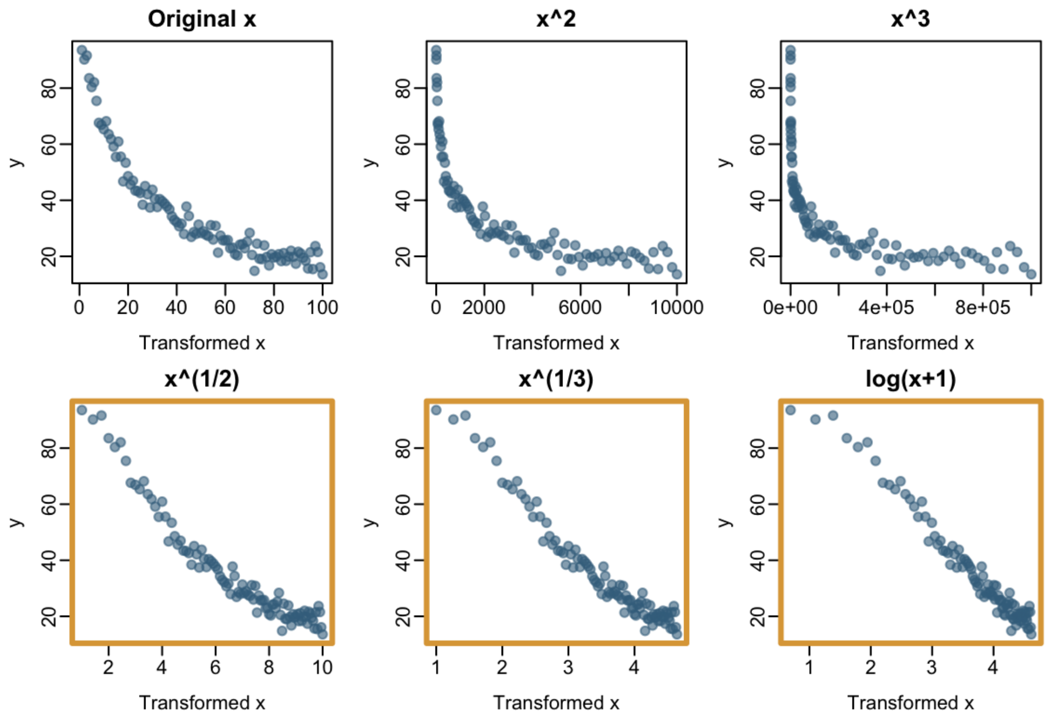
Note. As all original responses are 0 (cue not visited on a trial) or 1 (cue visited on a trial), raw data is visualized here with a cubic spline fit (λ = 0, i.e., no smoothing; JMP Student Edition 18.2.1). Error ribbons represent JMP's bootstrapped confidence region for the spline fits.

Effort Transform Tests

To address the potential of the shape of effort's effect to differ from that seen in delay discounting studies and be concave rather than convex (Białaszek et al., 2017; Klein-Flügge et al., 2015; Mitchell, 2017; see Figure 2), multiple transformations of the Effort variable were tested to be used in analyses. Visualizations of the tested transforms can be seen in Figures 10 and 11. If the shape of effort's effect is convex, like a traditional delay discounting curve, I would anticipate the best fitting transforms to be the traditional log transform, a square root or a cube root transform (see Figure 10). If the shape of effort's effect is concave like that seen in some effort discounting studies, I would anticipate the best fitting transforms to be the squared or cubed transforms (see Figure 11).

Figure 10

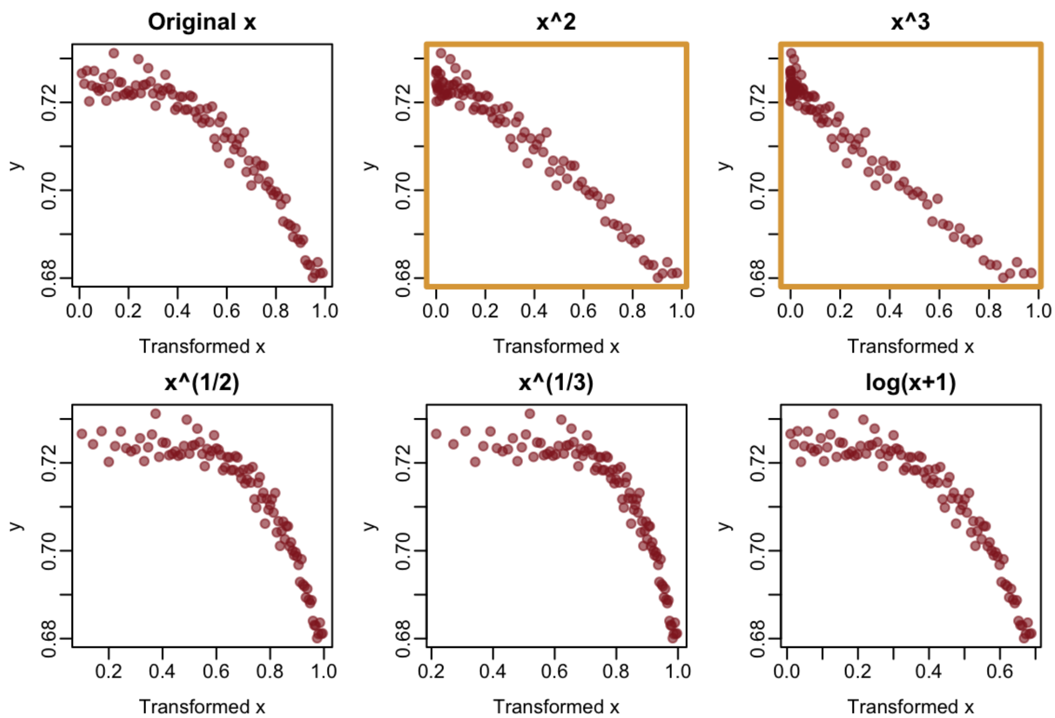
Visualization of Transforms of Simulated Convex Data



Note. Simulated data depicting a convex shape seen in traditional delay discounting curves, along with the transformations tested in this study. Here, the convex shape is best linearized by a square root, cube root, or log(x+1) transform.

Figure 11

Visualization of Transforms of Simulated Concave Data



Note. Simulated data depicting a concave shape seen in some effort discounting curves, along with the transformations tested in this study. Here, the concave shape is best linearized by a squared or cubed transform.

Transformations were tested on Experiment 1's data with a multilevel logistic regression, with an intercept effect of participant (R syntax provided in Equation 2). The model was run multiple times with different transformations of the effort variable, and the models are compared in Table 2. The log and cube root transforms were close in their AIC values, so I decided to use the more traditional log transform throughout this study. Importantly, the values of all tested transforms and the final log transform were multiplied by −1 before their use in further analyses. This is due to the combined effort term (mouse speed coefficient / distance in proportion of screen height) scaling such that higher effort is represented by lower values. In order to ease the interpretation of this value, it is multiplied by −1 such that higher effort is represented by higher values. The transformed Effort and Predictive Validity variables were centered by their minimums, and the variable of Trial was rescaled between 0 and 1 (dividing by the total trial number of 270) to ease in model convergence.

$$\begin{gathered}\text{Model} <\!- \ \text{glmer}(\text{visited} \sim -1(\text{EFFORT}) - \min(-1(\text{EFFORT})) \times \\ (\log(PV+1) - \min(\log(PV+1) \times (\text{Trial} / 270) + \\ + (1 \mid \text{Participant})\end{gathered} \tag{2}$$

Table 2

Model Comparisons of Effort Transforms in Logistic Regression Predicting Likelihood of Cue Visits for Experiment 1

Transform of Combined Effort Variable (x)	Formula of Transform	AIC
1. Cube Root	$x^{1/3}$	19162.14
2. Log	log(x + 1)	19164.27
3. Square Root	$x^{1/2}$	19177.18
4. No Transform	−x	19236.46
5. Squared	x^2	19350.02
6. Cubed	x^3	19446.39

Note. Model comparisons for transforms of the combined effort variable (mouse speed coefficient / distance in proportion of screen height) predicting likelihood of cue visits. All models compared had identical fixed and random effects, apart from the tested transformed effort variable (including fixed effects of minimum-centered log(predictive validity), Trial rescaled from 0 to 1, and their interaction). All models had an intercept random effect for participant. All transformed effort variables were multiplied by −1 after their transform, for ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). All transformed effort variables were additionally minimum-centered, to aid in interpretation of the model intercept (intercept represents all values at their minimum). Squared and cubed transforms were divided by their maximum values to rescale them between 0 and 1 to help with model convergence. The models with the lowest AIC values, cube root transform and log transform, were similar in their AIC value (19162.14 vs 19164.27). As these models did not significantly differ, the log transform was chosen for further models due to it being traditionally used with discounting models (Wileyto et al., 2004; Young, 2018).

Model Comparison

Of the original frequentist models run, the model with all two-way interactions in the random effects structure was the best of those that converged (see AIC comparison table in Appendix E). The maximal model, with two- and three-way interactions in the random effects structure, did not converge with bobyqa, nloptr, or Nelder-Mead optimizers. This model was then run as a Bayesian analysis, along with rerunning the best fit of the frequentist models (up to two-way interactions in random slope effects) for comparison. The WAIC values for this model comparison are shown in Table 3. As the maximal model that only converged through Bayesian methods was the best fitting model, the models reported for this experiment and the next will be Bayesian. The details of the best fitting frequentist models will be available in Appendix E, but the conclusions and interpretations ultimately do not differ between the two.

Table 3

WAIC Comparisons for Logistic Regressions Predicting Likelihood of Cue Visits for Experiment 1

Model	Random Effects Syntax	WAIC
1. Only Two-Way Interactions	(1 + Effort + Predictive Validity + Trial + Effort : Predictive Validity + Effort : Trial + Predictive Validity : Trial | Participant)	14221.5
2. Two- and Three-Way Interactions	(1 + Effort × Predictive Validity × Trial | Participant)	14101.6

Note. Model comparisons for Bayesian multilevel logistic regressions predicting likelihood of cue visits. Models only differed in their random effects. The fixed effects Effort, Predictive Validity, Trial, and their interactions, were kept the same for all models. The parameter of "Effort" is made up of −1(log[Effort] + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of Trial is made up of trial number / 270. The R syntax for each random effect structure is included. With the lowest WAIC value, Model 2 was chosen as the best fit.

Model Parameter Estimates

All compared Bayesian models were run with 8000 total iterations, with 2000 burn-in iterations. Priors for this model included a prior for the intercept, fixed effects, and random effects following a student's *t* distribution, with 3 degrees of freedom, 0 as the mean, and a scale of 10. This prior was made to be weakly informative, not restricting the parameter estimates because large parameter estimates were seen in the frequentist models run previously (see Appendix E).The prior for random effect correlations followed a LKJ(2) distribution. All R-hat values (< 1.01) posterior distributions and sampling chains suggest a converging model with no issues regarding parameter estimates. The parameter estimates for the best fitting Bayesian model are shown in Table 4. A visualization of the model-predicted conditional effects can be seen in Figure 12, and with effort back-transformed in Figure 13.

Table 4

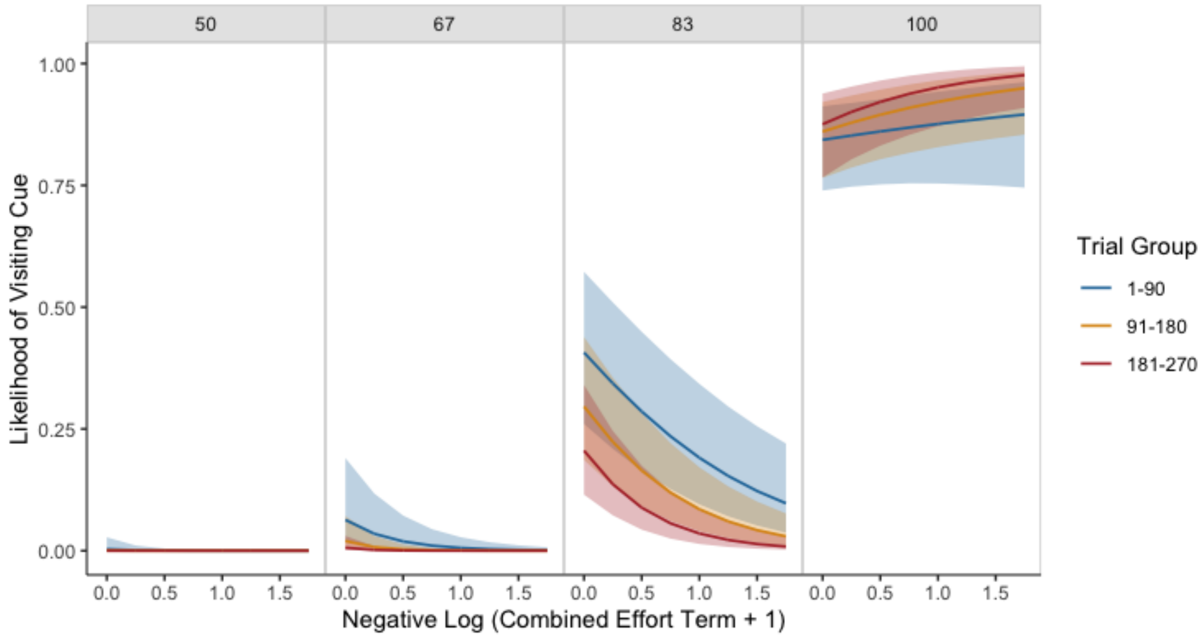
Parameter Estimates for Bayesian Logistic Regression Predicting Likelihood of Cue Visits for Experiment 1

	B	L−95% CI	U−95% CI	*BF*
Intercept	−4.87	−7.26	−2.61	>1000
Effort	−3.44	−5.06	−2.03	>1000
Predictive Validity	9.50	5.91	13.33	>1000
Trial	−6.51	−10.13	−3.06	>1000
Effort × Predictive Validity	5.15	2.80	7.88	>1000
Effort × Trial	−8.03	−12.22	−4.16	>1000
Predictive Validity × Trial	10.11	4.35	16.22	>1000
Effort × Predictive Validity × Trial	13.40	7.05	20.28	>1000

Note. Parameter estimates for the Bayesian multilevel logistic regression predicting likelihood of cue visits. The parameter of “Effort” is made up of −1(log(Effort+ 1) centered by its minimum. The parameter of “Predictive Validity” is made up of log(predictive validity + 1) centered by its minimum. The parameter of Trial is made up of trial number / 270. This model included random intercept and slope effects of effort, predictive validity, trial, and all possible interactions. Estimates are given in log odds space. Bayes Factors are the evidence in favor of the sign of the parameter estimate.

Figure 12

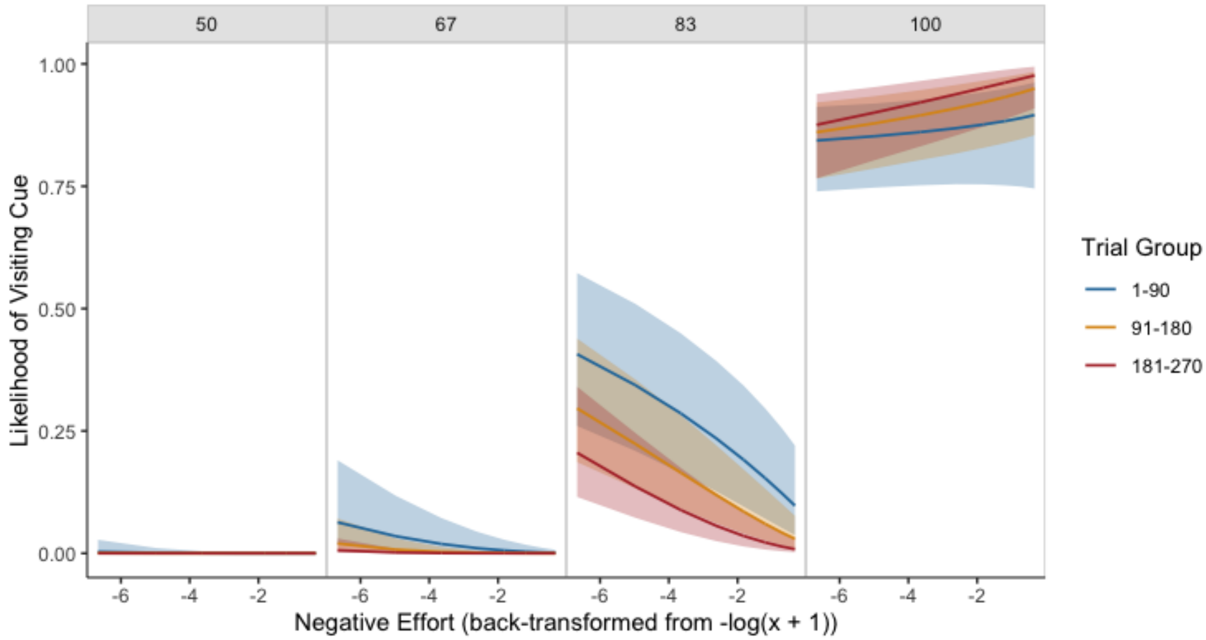
Conditional Effects of Logistic Regression for Experiment 1



Note. In this figure, the Effort variable appears as it did in the analyses, log transformed (log(x+1)), minimum-centered, and multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Credible Intervals.

Figure 13

Conditional Effects of Logistic Regression for Experiment 1, with Effort Back-Transformed



Note. In this figure, the Effort variable is back-transformed to its original scale but still multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Credible Intervals.

This model shows support for both Hypotheses 1A and 1B under different conditions. Effort has the greatest effect for the second highest predictive validity value, with participants visiting cues 29.5% of the time on average when effort is low (the conditional effect of the model when the Effort variable (minimum centered −1(log(Effort + 1)) is set to 0*,* Predictive Validity is set to 87, and all other variables are set to their means; 95% CI = [0.18, 0.42]) and 2.8% of the time when effort is high (at value of 1.75 for minimum centered −1(log(Effort + 1)); 95% CI = [0.01, 0.06]*)*.This difference is smaller for the 100% correct cue (*M* = 0.86, 95% CI = [0.78, 0.93] at low effort to $M = 0.95$, 95% CI = [0.88, 0.99] at high effort) and smaller still for the 67% ($M = 0.02$, 95% CI = [0.002, 0.06] at low effort to $M = 1.63e^{-5}$, 95% CI = [$1.00e^{-7}$, $1.24\ e^{-4}$] at high effort) and 50% cues ($M = 2.93\ e^{-4}$, 95% CI = [$5.00e^{-7}$, 0.002] at low effort to $M = 0$, 95% CI = [0, 0] at high effort).

Overall, the largest visible effect is that of Predictive Validity, particularly in the leap from the 83% to the 100% correct cue. This leap is strong enough that it prompted visual comparison back to the raw data (Figure 9). Fitting what would be expected from Prospect Theory and the potential for a certainty effect (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), the leap from 83% to 100% Predictive Validity may be a larger conceptual leap than the differences between any other two sequential Predictive Validity values. This difference may be enough such that the 100% cue single-handedly drives a significant effect of Predictive Validity, dragging up the estimated visit likelihoods for the 83% cue in the process (compare the higher predicted values for the 83% cue in Figure 13 to the raw data in Figure 9). To investigate this further, the model was re-run with Predictive Validity as a categorical variable.

Model Comparison with Predictive Validity as a Categorical Variable

Two models were compared with Predictive Validity (log[predictive validity + 1] centered by its minimum) as a Categorical Variable (dummy coded with zero representing the minimum value) One of the compared models had only up to two-way interactions as slope effects in the random effects structure, while the other had up to the three-way interaction of Effort, Predictive Validity, and Trial. Of these, the model with only up to the two-way interactions between these variables was the best fit and used for further analyses. See details of this comparison in Table 5. The model with only two-way interactions in the random effects structure was also the best fitting of the frequentist models run. For more details on this model, see Appendix F. Importantly, the WAIC values for the models using Predictive Validity as a categorical variable were lower than the best-fitting model with Predictive Validity as a continuous variable (WAIC of 10072.2 for best fitting model with categorical Predictive Validity, WAIC of 14101.6 for best fitting model with continuous Predictive Validity), suggesting the categorical treatment of Predictive Validity is the best fit.

Table 5

WAIC Comparisons for Bayesian Logistic Regressions with Predictive Validity as a Categorical Variable Predicting Likelihood of Cue Visits for Experiment 1

Model	Random Effects Syntax	WAIC
1. Only Two-Way Interactions	(1 + Effort + Predictive Validity + Trial + Effort : Predictive Validity + Effort : Trial + Predictive Validity : Trial | Participant)	10072.2
2. Two- and Three-Way Interactions	(1 + Effort × Predictive Validity × Trial | Participant)	10077.4

Note. Model comparisons for Bayesian multilevel logistic regressions predicting likelihood of cue visits. Models only differed in their random effects. The fixed effects Effort, Predictive Validity, Trial, and their interactions, were kept the same for all models. The parameter of "Effort" is made up of −1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of Trial is made up of trial number / 270. The R syntax for each random effect structure is included. With the lowest WAIC value, Model 1 was chosen as the best fit. For comparison, the best fitting model with Predictive Validity as a continuous predictor had an WAIC of 14101.6

Parameter Estimates and Conditional Effects of the Best Fitting Model

The parameter estimates for the best fitting model, with Predictive Validity as a categorical variable, can be seen in Table 6. Conditional Effects from this model are visibly closer to the raw data in Figure 9 (see Figure 14 for model conditional effects, and Figure 15 for conditional effects with effort back-transformed). The better fitting model using Predictive

Validity as a categorical variable shows strong support for Hypothesis 1B. In the case of this study, 100% correct information drives cue visits, even when it requires increased effort.

Table 6

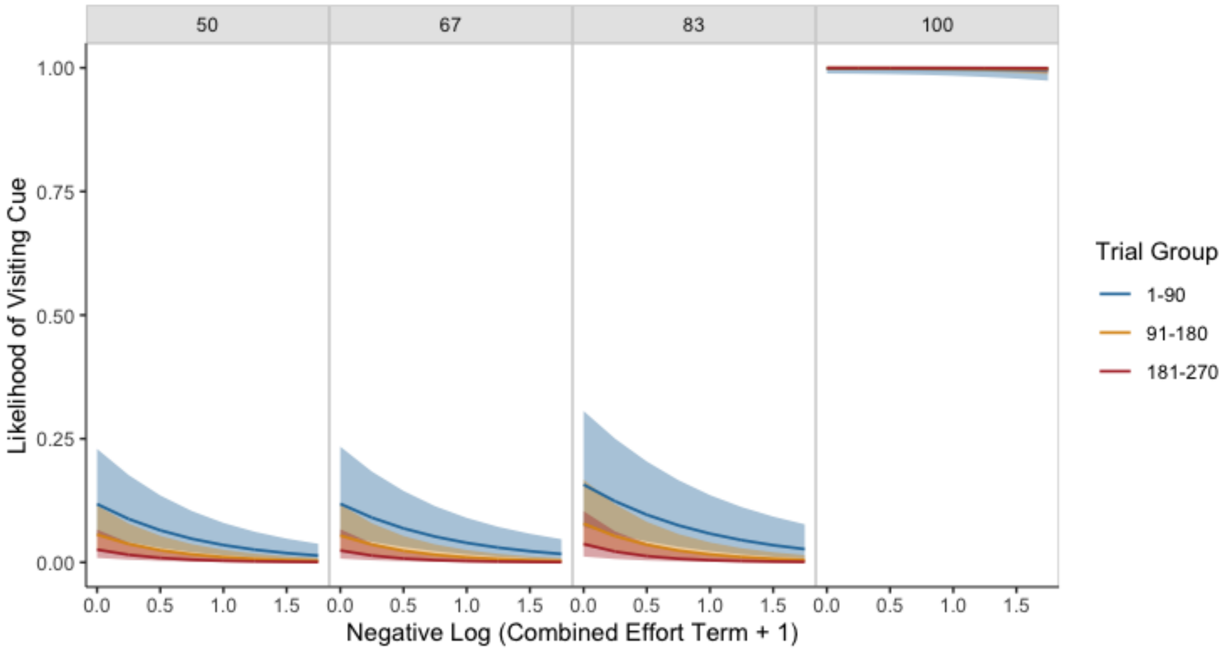
Parameter Estimates for Bayesian Logistic Regression with Predictive Validity as a Categorical Variable Predicting Likelihood of Cue Visits for Experiment 1.

	B	L−95% CI	U−95% CI	*BF*
Intercept	−1.62	−2.53	−0.73	>1000
Effort	−1.13	−1.65	−0.63	>1000
Predictive Validity [67]	0.02	−0.37	0.41	1.21
Predictive Validity [83]	0.33	−0.13	0.80	11.67
Predictive Validity [100]	7.27	5.27	9.86	>1000
Trial	−2.45	−3.85	−1.09	>1000
Effort: Predictive Validity [67]	0.17	−0.24	0.58	3.85
Effort: Predictive Validity [83]	0.25	−0.24	0.73	5.53
Effort: Predictive Validity [100]	0.99	−0.08	2.15	28.52
Effort:Trial	−1.24	−2.36	−0.19	99.00
Predictive Validity [67]:Trial	−0.12	−1.11	0.78	1.44
Predictive Validity [83]:Trial	0.06	−0.96	1.00	1.26
Predictive Validity [100]:Trial	7.53	3.65	12.05	>1000
Effort:Predictive Validity [67]:Trial	−0.40	−1.28	0.47	4.43
Effort: Predictive Validity [83]:Trial	−0.39	−1.27	0.50	4.09
Effort: Predictive Validity [100]:Trial	−0.09	−1.64	1.47	1.19

Note. Parameter estimates for the Bayesian multilevel logistic regression predicting likelihood of cue visits. The parameter of "Effort" is made up of −1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The labels of the Predictive Validity levels are written in their original scale (67, 83, 100). The categorical variable of Predictive Validity is dummy coded, with the minimum level 50 as the reference level. The parameter of Trial is made up of trial number / 270. This model included random intercept and slope effects of effort, predictive validity, trial, and all possible two-way interactions. Estimates are given in log odds space. Bayes Factors are evidence in favor of the sign of the parameter estimate.

Figure 14

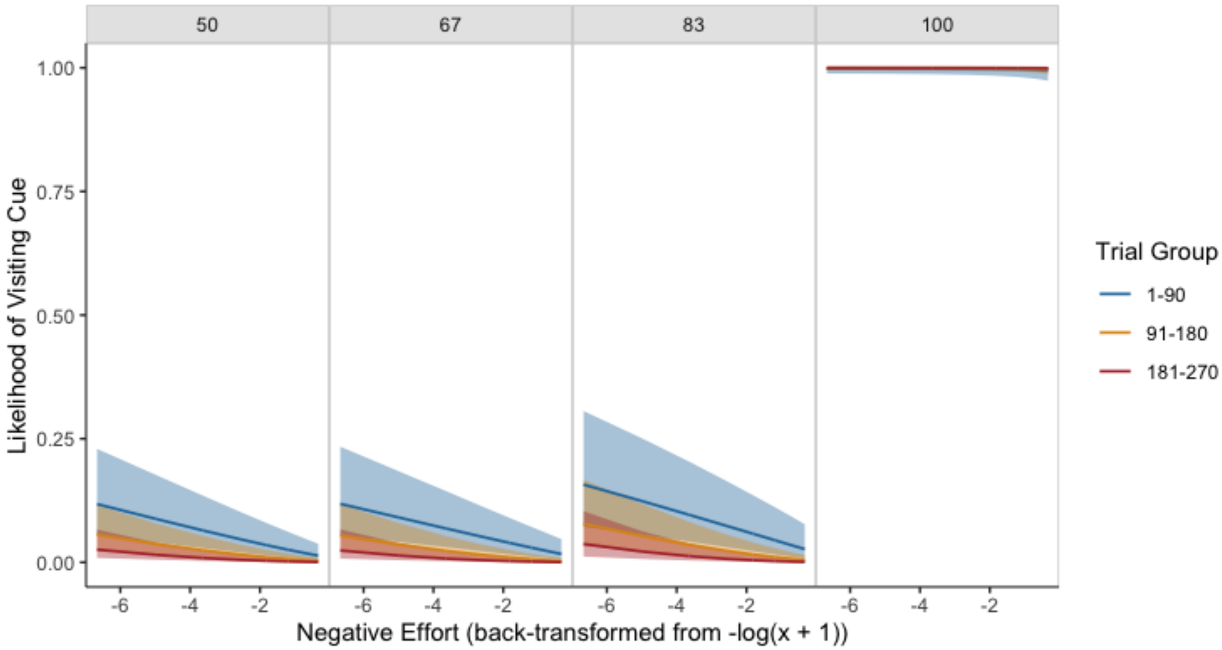
Conditional Effects of Logistic Regression for Experiment 1, with PV as a Categorical Predictor



Note. In this figure, the Effort variable appears as it did in the analyses, log transformed (log(x+1)), minimum-centered, and multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Credible Intervals.

Figure 15

Conditional Effects of Logistic Regression for Experiment 1, Categorical PV, Back-Transformed



Note. In this figure, the Effort variable is back-transformed to its original scale but still multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Credible Intervals.

This model shows decisive evidence in favor of an effect of Effort ($B = -1.13$, 95% CI = [−1.65, −0.63], $BF > 1000$), showing support for Hypothesis 1A. The largest effect seen is that of Predictive Validity at the conditional level of 100 ($B = 7.27$, 95% CI = [5.27, 9.86], $BF > 1000$), and its interaction with Trial ($B = 7.53$, 95% CI = [3.65, 12.05], $BF > 1000$). Where other Predictive Validity values show a decrease in cue visit likelihood over the course of the experiment (for Predictive Validity 50%, at Trial 1, $M = 0.07$, 95% CI = [0.02, 0.14] and at Trial 270, $M = 0.002$, 95% CI = [0.0002, 0.006]; For 67% at Trial 1 $M = 0.08$, 95% CI = [0.02, 0.16] and at Trial 270, $M = 0.005$, 95% CI = [0.0001, 0.005]; and for 83% at Trial 1 $M = 0.11$, 95% CI = [0.03, 0.23] and at Trial 270, $M = 0.003$, 95% CI = [0.0002, 0.008]), the 100% cue is relatively unchanged (at Trial 1 $M = 0.99$, 95% CI = [0.98, 0.99] and at Trial 270, $M = 0.99$, 95% CI = [0.99, 1.00]).

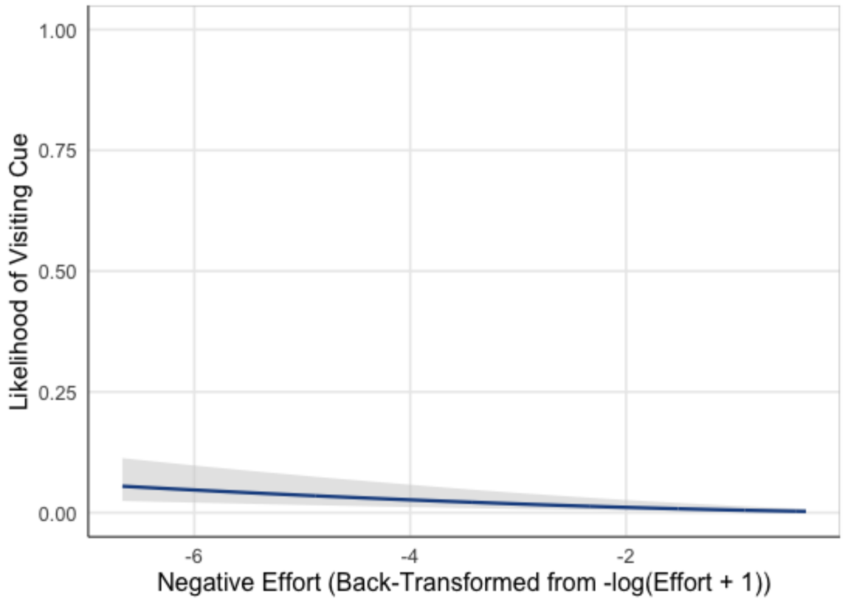
Both Hypothesis 1A, that effort can discount the value of information and inhibit the search for cues, and Hypothesis 1B, that a need for certainty outweighs the effort required to get the cues, appear to be supported under different conditions, characterized by significant interactions of Effort and the 100% Predictive Validity cue ($B = 0.99$, 95% CI = [−0.08, 2.15], $BF = 28.52$). Visits to cues with predictive validities of 50, 67, and 83 appear to be strongly influenced by effort (50 PV: likelihood of cue visit at low effort of 0 for minimum centered −1(log(Effort + 1)), $M = 0.05$, 95% CI = [0.02, 0.10], higher than at high effort of 1.75 $M = 0.003$, 95% CI = [0.0004, 0.007]; $OR = 21.35$, 95% CI = [6.67,45.00]; 67 PV: likelihood of cue visit at low effort, $M = 0.05$, 95% CI = [0.02, 0.10], higher than at high effort, $M = 0.002$, 95% CI = [0.0004, 0.006], $OR = 22.52$, 95% CI = [7.842,49.5]; 83 PV: likelihood of cue visit at low effort , $M = 0.08$, 95% CI = [0.03, 0.15], higher than at high effort $M = 0.004$, 95% CI = [0.001, 0.01]), $OR = 19.36$, 95% CI = [6.298,42.1]), whereas the cue with predictive validity of 100 was not (100 PV: likelihood of cue visit at low effort, $M = 0.99$, 95% CI = [0.99, 1.00], not different than at high effort 1.75, $M = 0.99$, 95% CI = [0.99, 0.99], $OR = 4.24$, 95% CI = [0.219;15.2]). A visualization of the likelihood of visiting the 50, 67, or 83 cue on trials where the 100 cue is not visited can be seen in Appendix G (609 instances, out of 8,100 total [30 participants × 270 trials]).

Predictive validity does not have a large effect on cue visits between the 50, 67, and 83 cues (for 50, with all other variables at their average value, M likelihood of cue visits = 0.01, 95% CI [0.004,0.03]; for 67, $M = 0.01$, 95% CI [0.003,0.02]; for 83, $M = 0.02$, 95% CI [0.004,0.04]), but the likelihood of visiting the 100 cue is much larger than all others ($M = 0.99$, 95% CI [0.99,1.00], compared to 83, $OR = 89906$, 95% CI [675, 1,015,748]. As such, both Hypotheses 1A, the Effort Dominance Hypothesis, and 1B, the Predictive Validity Dominance Hypothesis, are supported under different conditions: effort has a strong effect with non-100% Predictive Validity cues, and predictive validity has a strong effect only with the 100% Predictive Validity cue.

Experiment 1 suggests the shape of the function of cue visits predicted by effort is convex, as predicted by Hypothesis 3B, the Largest Discounting Early Hypothesis. This is supported by the best fitting transform for the effort variable, a log transform (AIC = 19164.27; though cube root is also tied for best fit, AIC = 19162.14). A visualization of the conditional effects of the best fitting model with Predictive Validity as a categorical predictor, with the effort predictor back transformed (yet still multiplied by −1 to aid in interpretation, with higher values meaning more effort), shows no strong pattern of concave or convex shapes of the effort effect (see Figure 15). When examining the conditional effect of effort when all other variables are set to their means (see Figure 16), a very slight concave shape appears (slope changing from −0.012 at the low effort value of −6 to a flatter slope of −0.006 at the high effort value of −2). For the best fitting model with Predictive Validity as a continuous predictor (see Figure 17), these slopes are only slightly different at −0.012 at the low effort value of −6 to −0.003 at the high effort value of −2, still suggesting a slight convex shape. In sum, these results tilt in favor of Hypothesis 3B, but not to a powerful degree.

Figure 16

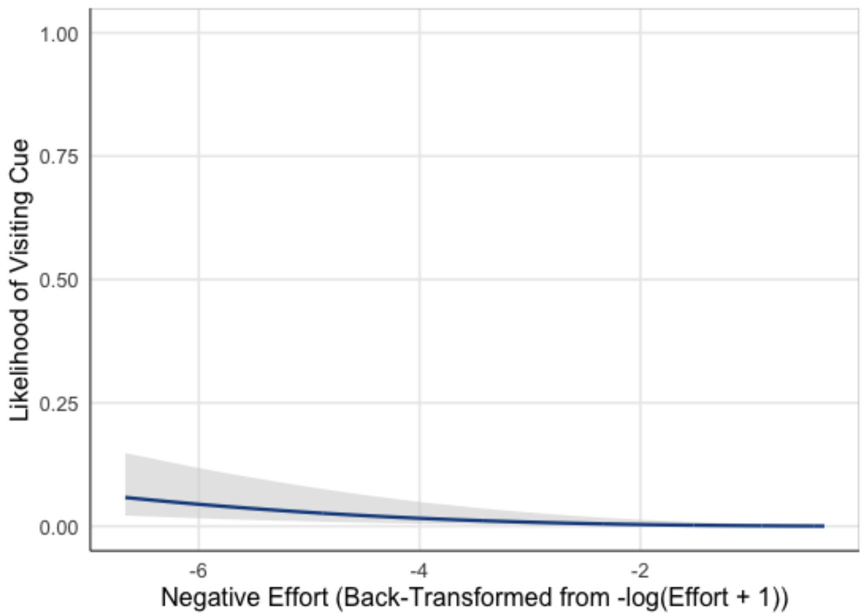
Conditional Effect of Effort for Logistic Regression with Categorical PV, Experiment 1



Note. Here, from the multilevel logistic regression predicting likelihood of cue visits, with Predictive Validity as a categorical predictor, the conditional effect of Effort is shown when all other variables are at their means. The Effort variable is back-transformed to its original scale but still multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Error ribbon represents 95% Credible Interval.

Figure 17

Conditional Effect of Effort for Logistic Regression with Continuous PV, Experiment 1



Note. Here, from the multilevel logistic regression predicting likelihood of cue visits, the conditional effect of Effort is shown when all other variables are at their means. The Effort variable is back-transformed to its original scale but still multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Error ribbon represents 95% Credible Interval.

Experiment 2

Sample Demographics

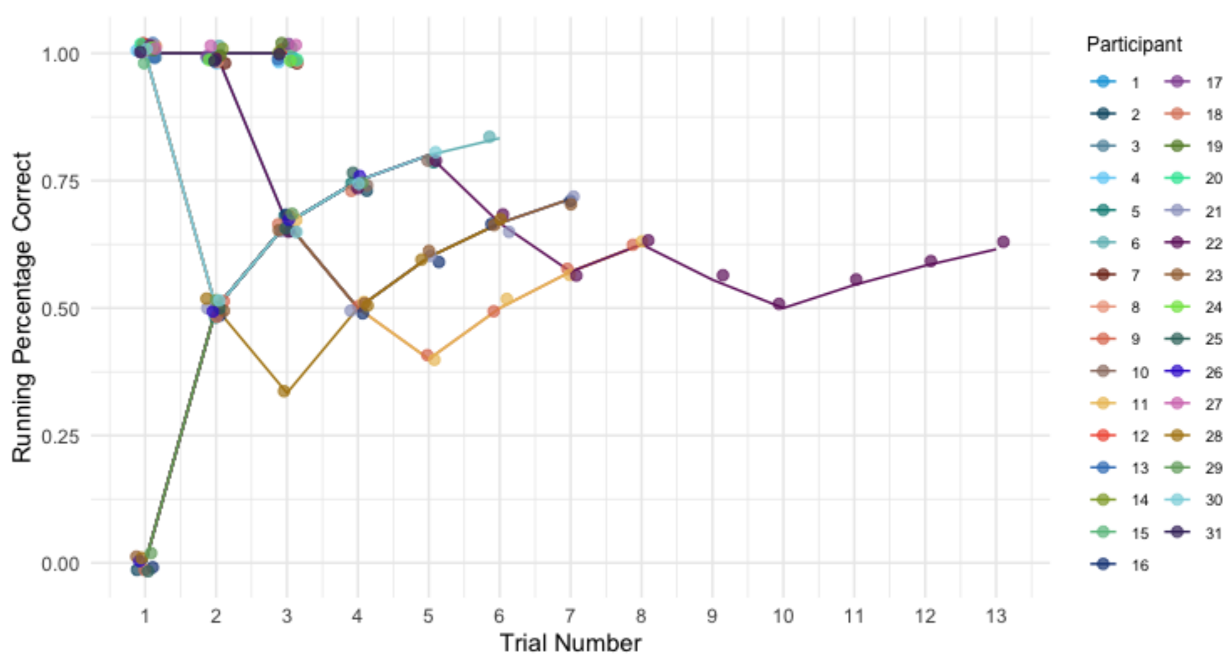
Experiment 2 originally recruited 35 participants. Of these, one was removed for reporting a technical issue, and three were removed for exiting the full-screen window during the study (reported screen size decreased partway during the experiment). Subjects were recruited until a final sample of 31 was reached. The sample for Experiment 2 had 16 female and 15 male participants, with a mean age of 43 (*SD* = 12).

Raw Data Visualizations and Descriptive Statistics

Participants in Experiment 2 started each experiment block of 45 trials with exploration of multiple cues but increasingly visited and relied only on the 100% cue over the course of each block. Most participants completed the practice round requiring three correct answers in a row within four trials (23/31 participants; see Figure 18), and all participants finished the practice within 13 trials. No participant got through the practice round by solely guessing and visiting no cues (see number of cue visits per trial for the practice round in Figure 19).

Figure 18

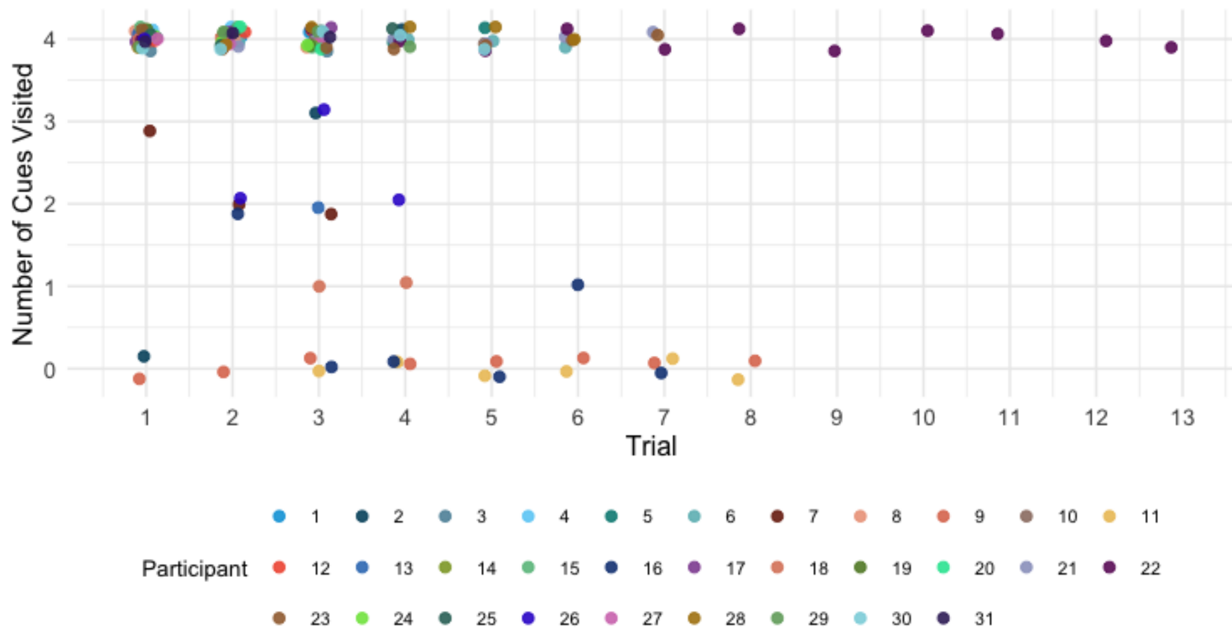
Running Proportion of Correct Responses from Participants in Practice Trials of Experiment 2



Note. The line representing each participant's running proportion of correct responses will deflect downwards when that participant makes an incorrect response.

Figure 19

Number of Cues Visited by Participants in Practice Trials of Experiment 2



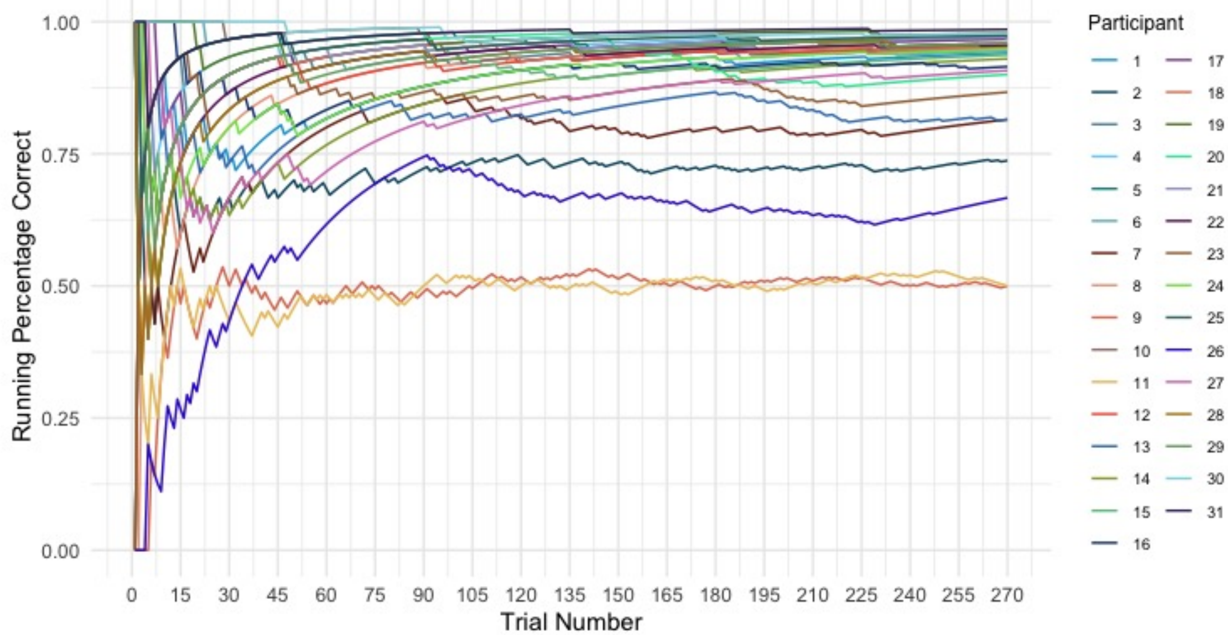
Note. Number of cues visited by each participant in the practice trials for Experiment 2. Practice trials continue until the participant gets three responses correct in a row.

In the full study, most participants were able to maintain a running percentage of correct responses of 75% or higher by the end of Block 1 (24 of 31 participants, see Figure 20). Only four participants seemed to have exhibited a strategy of guessing without cue visits, resulting in a running percentage of correct responses that centered around or drifted towards 50% (chance; see Figure 20). These four participants are also visible as frequently visiting 0 in a visualization of the number of cue visits per trial, Figure 21. Figure 21 also shows that the large majority of subjects began to rely on only one cue as each experiment block progressed. Similar to as was found in Experiment 1, visiting one cue per trial led to the fastest trial plus ITI duration (*Mdn* = 4.39 s, *IQR* = [3.66, 5.65]), suggesting that the one-cue-visit strategy (and visiting only the 100%

cue) minimized experiment duration. The average duration of each trial plus potential penalty ITI is plotted in Appendix I, by the number of cues visited on each trial.

Figure 20

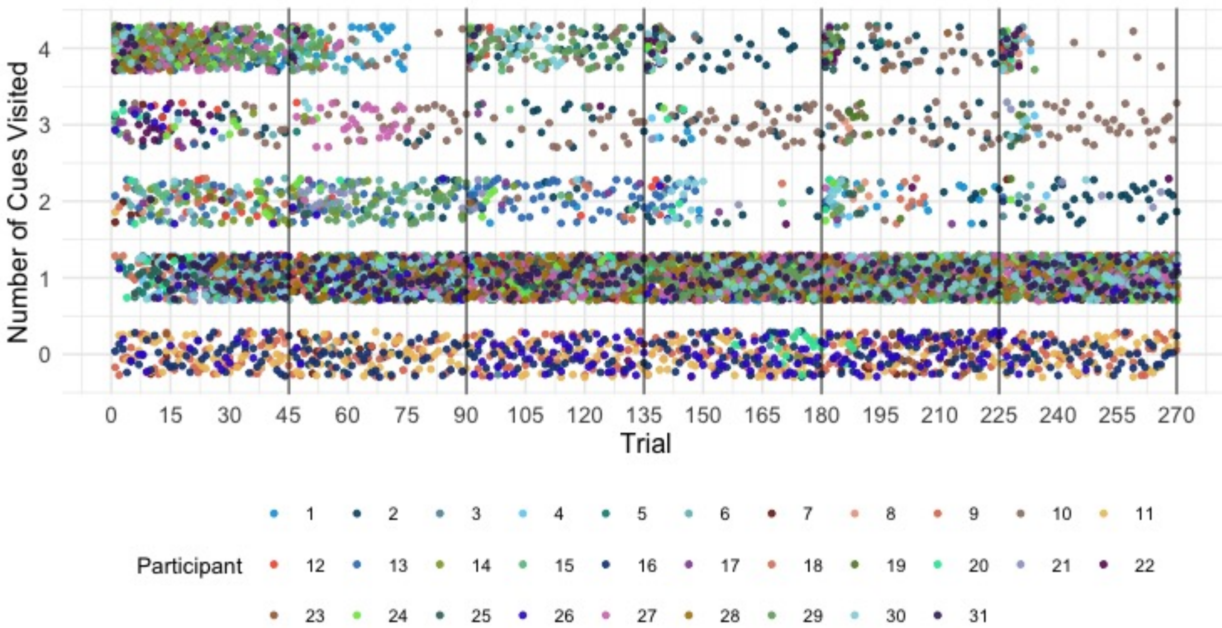
Running Proportion of Correct Responses from Participants in Experiment 2



Note. The line representing each participant's running proportion of correct responses will deflect downwards when that participant makes an incorrect response.

Figure 21

Number of Cues Visited by Participants in Experiment 2



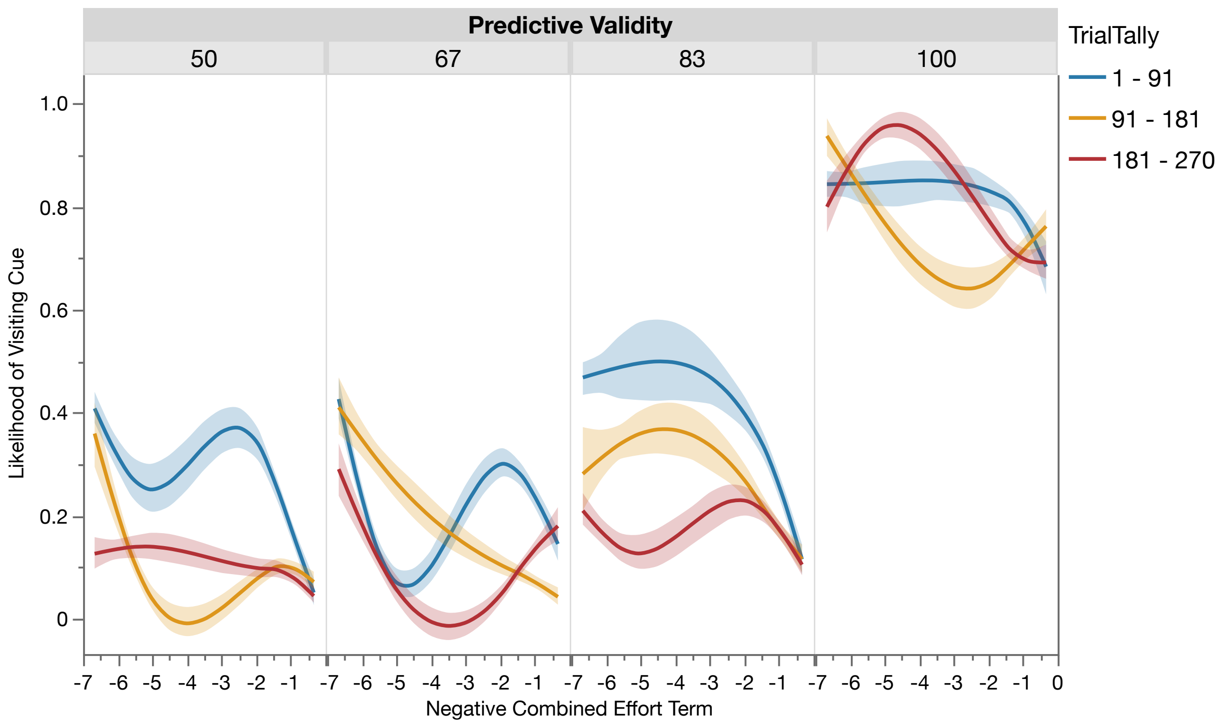
Note. Number of cues visited by each participant in each block of Experiment 2, by trial. Each block of 45 trials is indicated by a black line. By the end of the trials in a block, participants gravitate towards viewing only one cue.

Participants also moved their mouse cursors with greater physical movements when the mouse speeds were manipulated to be slower. Visualization of the outcome mouse movement speeds under the different speed coefficient settings for Experiment 2 can be seen in Appendix J. When the mouse speed was set to 80% of its original speed, participants moved at a median speed of 0.43 screen height units / s (*IQR* = [0.25 , 0.67]). This was 88% of their median for 100% speed (*Mdn* = 0.49 screen height / s, *IQR* = [0.28, 0.73]), achievable by moving the mouse or trackpad physically more than was done at the 100% speed setting. Participants made even larger physical movements to compensate for the slowest mouse speed of 20%, with participants moving a median of 0.19 screen height / s (*IQR* = [0.11 , 0.35]), which was 39% of the median for 100% speed.

The raw data is visualized in Figure 22, separated by the predictor variables of interest. Similar to Experiment 1, the raw data for Experiment 2 immediately shows a distinct difference in responses for the 100% cue compared to the other cues. Additional visualizations of this raw data, including data separated by Participant and data separated by the individual effort factors of mouse speed and distance, can be seen in Appendix H.

Figure 22

Raw Data for Experiment 2 by Effort, Predictive Validity, and Trial



Note. As all original responses are 0 (cue not visited on a trial) or 1 (cue visited on a trial), raw data is visualized here with a cubic spline fit (λ = 7 to reduce excessive curvature and predictions above 1 and below 0; JMP Student Edition 18.2.1). Error ribbons represent JMP's bootstrapped confidence region for the spline fits.

Effort Transform Tests

As was done in Experiment 1, different transforms of the Effort variable were tested to find the best fit. Visualizations of the tested transforms can be seen in Figures 10 and 11. A convex functional form, like a traditional delay discounting curve, would be best fit by the traditional log transform, a square root or a cube root transform (see Figure 10). A concave functional form, like that seen in some effort discounting studies, would be best fit by squared or cubed transforms (see Figure 11).

Transformations were tested on Experiment 2's data with a multilevel logistic regression, with an intercept effect of participant (R syntax provided in Equation 3). The model was run multiple times with different transformations of the effort variable, and the models are compared in Table 7. The log and cube root transforms were close in their AIC values, though not as close as they were in Experiment 1 (for log, AIC = 26353.54. For cube root, AIC = 26342.04). As both transforms suggest a convex shape, and for consistency with Experiment 1 and traditional transforms of delay discounting data (Wileyto et al., 2004; Young, 2018), I decided to use the more traditional log transform throughout this study. Importantly, the values of all tested transforms and the final log transform were multiplied by −1 before their use in further analyses. This is due to the combined effort term (mouse speed coefficient / distance in proportion of screen height) scaling such that higher effort is represented by lower values. In order to ease the interpretation of this value, it is multiplied by −1 such that higher effort is represented by higher values. The transformed Effort and Predictive Validity variables were centered by their minimums, and the variable of Trial was rescaled between 0 and 1 (dividing by the total trial number of 270) to ease in model convergence.

$$\begin{gathered}\text{Model} <\!- \text{glmer(visited} \sim -1(\text{EFFORT}) - \min(-1(\text{EFFORT})) \times \\ (\log(\text{PV}+1) - \min(\log(\text{PV}+1) \times (\text{Trial} / 270) \times \text{Block} + \\ + (1 \mid \text{Participant})\end{gathered} \tag{3}$$

Table 7

Model Comparisons for Effort Transforms in Logistic Regression Predicting Likelihood of Cue Visits in Experiment 2

Transform of Combined Effort Variable (x)	Formula of Transform	AIC
1. Cube Root	$x^{1/3}$	26342.04
2. Log	log(x + 1)	26353.54
3. Square Root	$x^{1/2}$	26368.92
4. No Transform	-x	26435.66
5. Squared	x^2	26506.20
6. Cubed	x^3	26541.95

Note. Model comparisons for transforms of the combined effort variable (mouse speed coefficient / distance in proportion of screen height) predicting likelihood of cue visits. All models compared had identical fixed and random effects, apart from the tested transformed effort variable (including fixed effects of minimum-centered log(predictive validity), Trial rescaled from 0 to 1, and their interaction). All models had an intercept random effect for participant. All transformed effort variables were multiplied by −1 after their transform, for ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). All transformed effort variables were additionally minimum-centered, to aid in interpretation of the model intercept (intercept represents all values at their minimum). Squared and Cubed transforms were divided by their maximum values to rescale them between 0 and 1 to help with model convergence. To stay consistent with Experiment 1 and traditional transforms of delay discounting data (Wileyto et al., 2004; Young, 2018), the log transform was chosen (AIC = 26353.54) , even though it was only the second-best fitting transform after the less conventional cube root transform (AIC = 26342.04).

Model Comparison

The first models tested for Experiment 2 were those originally planned with Predictive Validity as a continuous predictor. These models will be examined first, followed by models where Predictive Validity is treated as a categorical variable, as was run for Experiment 1.

For consistency with the models reported for Experiment 1, and because many of the corresponding frequentist models for Experiment 2 failed to converge, the Bayesian models are thus reported for Experiment 2. The best-fitting converging frequentist model run with Predictive Validity as a continuous predictor can be viewed in Appendix K. No frequentist models with Predictive Validity as a categorical predictor converged. Of the Bayesian models run, the maximal model with all possible interactions of Effort, Predictive Validity, Trial, and Block as slope effects in the random effects structure, was the best fit (see Table 8).

Table 8

WAIC Comparisons for Logistic Regressions Predicting Likelihood of Cue Visits for Experiment 2

Model	Random Effects Syntax	WAIC
1. Intercept Only	(1 | Participant)	26196.7
2. Main Effects	(1 + Effort + Predictive Validity + Trial + Block | Participant)	22321.7
3. Only Two-Way Interactions	(1 + Effort + Predictive Validity + Trial + Block + Effort : Predictive Validity + Effort : Trial + Predictive Validity : Trial + Block : Effort + Block : Trial + Block : Predictive Validity | Participant)	18489.7
4. Two- and Three-Way Interactions	(1 + Effort × Predictive Validity × Trial × Block – Effort : Predictive Validity : Trial : Block | Participant)	16672.6
5.Two-, Three-, and Four-Way Interactions	(1 + Effort × Predictive Validity × Trial × Block | Participant)	16265.7

Note. Model comparisons for Bayesian multilevel logistic regressions predicting likelihood of cue visits. Models only differed in their random effects. The fixed effects Effort, Predictive Validity, Trial, and their interactions, were kept the same for all models. The parameter of "Effort" is made up of −1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of Trial is made up of trial number / 270. The R syntax for each random effect structure is included. With the lowest WAIC value, Model 5 was chosen as the best fit.

Similar to Experiment 1, all compared Bayesian models were run with 8000 total iterations, with 2000 burn-in iterations. Priors for this model included a prior for the intercept, fixed effects, and random effects following a student's *t* distribution, with 3 degrees of freedom, 0 as the mean, and a scale of 10. This prior was made to be weakly informative, not restricting the parameter estimates because large parameter estimates were seen in the frequentist models run previously (see Appendix E).The prior for random effect correlations followed a LKJ(2) distribution. All R-hat values (< 1.01) posterior distributions and sampling chains suggest a converging model with no issues regarding parameter estimates.

Model Parameter Estimates

The parameter estimates for the best fitting Bayesian model are shown in Table 9. A visualization of the model-predicted conditional effects can be seen in Figure 23, and with effort back-transformed in Figure 24.

Table 9

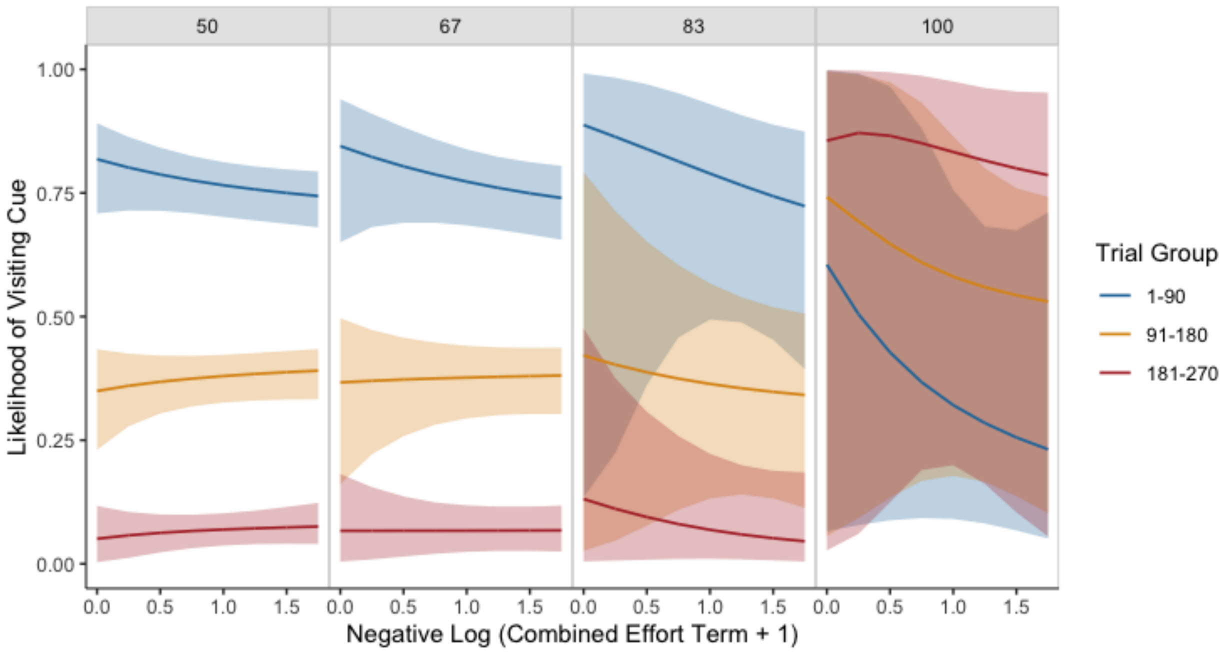
Parameter Estimates for Bayesian Logistic Regression Predicting Likelihood of Cue Visits for Experiment 2

	B	L−95% CI	U−95% CI	*BF*
Intercept	2.64	−2.83	8.05	4.68
Effort	−9.24	−13.25	−4.91	>1000
Predictive Validity	1.79	−4.79	8.40	2.41
Trial	−66.75	−83.40	−48.63	>1000
Block	4.79	2.23	7.20	>1000
Effort × Predictive Validity	11.68	4.68	17.89	>1000
Effort × Trial	−23.54	−49.87	−0.49	44.63
Predictive Validity × Trial	94.93	66.04	121.87	>1000
Effort × Block	5.89	3.00	8.97	>1000
Predictive Validity × Block	−8.55	−12.44	−4.65	>1000
Trial × Block	4.31	2.40	6.25	>1000
Effort × Predictive Validity × Trial	47.51	5.86	96.04	134.60
Effort × Predictive Validity × Block	−9.45	−14.89	−4.17	>1000
Effort × Trial × Block	−1.15	−3.32	1.10	5.77
Predictive Validity × Trial × Block	−4.70	−8.35	−0.86	113.83
Effort × Predictive Validity × Trial × Block	0.29	−4.04	4.50	1.23

Note. Parameter estimates for the Bayesian multilevel logistic regression predicting likelihood of cue visits for Experiment 2. The parameter of "Effort" is made up of −1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of Trial is made up of trial number / 270. This model included random intercept and slope effects of effort, predictive validity, trial, and all possible interactions. Estimates are given in log odds space. Bayes Factors are the evidence in favor of the sign of the parameter estimate.

Figure 23

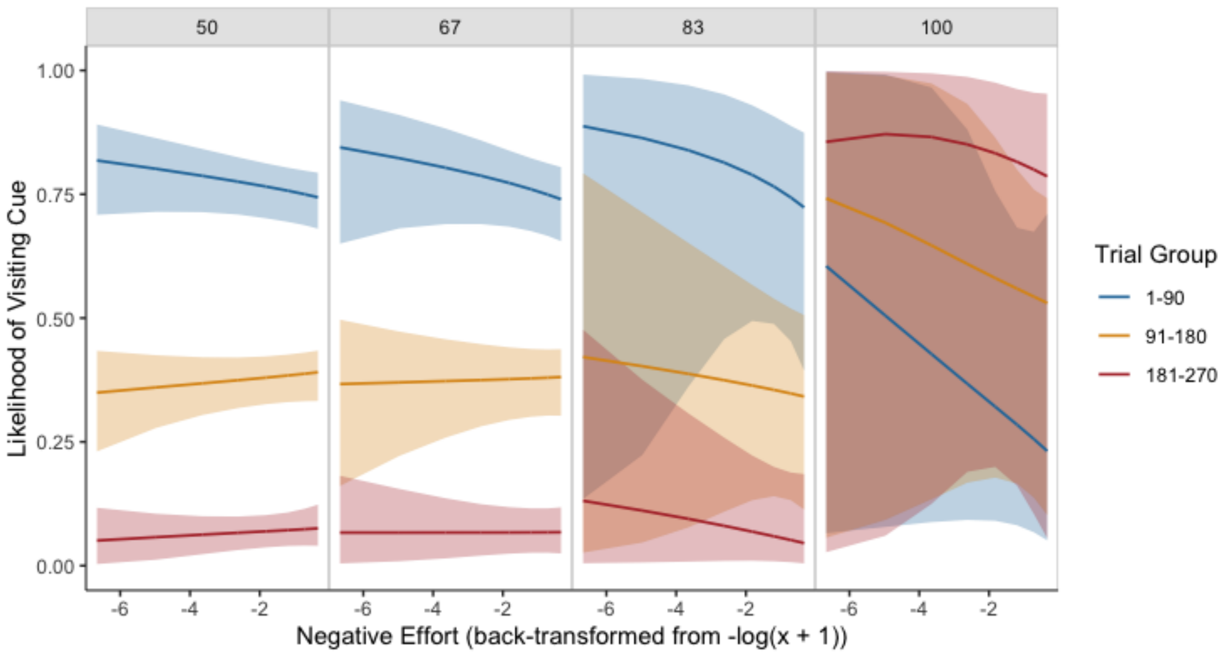
Conditional Effects of Logistic Regression for Experiment 2, with Predictive Validity as Continuous Variable



Note. In this figure, the Effort variable appears as it did in the analyses, log transformed (log(x+1)), minimum-centered, and multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Credible Intervals.

Figure 24

Conditional Effects of Logistic Regression for Experiment 2, with Predictive Validity as Continuous Variable and Effort Back-Transformed



Note. In this figure, the Effort variable is back-transformed to its original scale but still multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Credible Intervals.

Similar to Experiment 1, a visual inspection of the conditional effects from the best fitting model in Experiment 2 (see Figure 24) suggests the distinct nature of the 100% cue visible in the raw data (see Figure 22) may not be captured by treating Predictive Validity as a continuous variable. In particular, estimates for the 83% cue may be pulled upwards, and estimates for the 100% cue may be pulled down. As such, the data from Experiment 2 was similarly tested in models where Predictive Validity is treated as a categorical variable.

Model Comparison with Predictive Validity as a Categorical Variable

Five models with different random effects structures were compared with Predictive Validity as a Categorical Variable. Of these, the maximal model with all interactions up to the four-way interaction of Effort, Predictive Validity, Trial, and Block as slope effects in the random effects structure was the best fit and used for further analyses. See details of this comparison in Table 10. For details on the best-fitting frequentist model of those that converged, see Appendix K. Importantly, the WAIC values for the models using Predictive Validity as a categorical variable were lower than the best-fitting model with Predictive Validity as a continuous variable (WAIC of 11910.3 for best fitting model with categorical Predictive Validity, WAIC of 16265.7 for best fitting model with continuous Predictive Validity), suggesting the categorical treatment of Predictive Validity is the best fit

Table 10

WAIC Comparisons for Bayesian Logistic Regressions with Predictive Validity as a Categorical Variable Predicting Likelihood of Cue Visits for Experiment 2

Model	Random Effects Syntax	WAIC
1. Intercept Only	(1 | participant)	22986.8
2. Main Effects	(1 + Effort + Predictive Validity + Trial + Block | Participant)	17053.1
3. Only Two-Way Interactions	(1 + Effort + Predictive Validity + Trial + Block + Effort : Predictive Validity + Effort : Trial + Predictive Validity : Trial + Block : Effort + Block : Trial + Block : Predictive Validity | Participant)	14201.9
4. Two- and Three-Way Interactions	(1 + Effort × Predictive Validity × Trial × Block – Effort : Predictive Validity : Trial : Block | Participant)	12145.2
5.Two-, Three-, and Four-Way Interactions	(1 + Effort × Predictive Validity × Trial × Block | Participant)	11910.3

Note. Model comparisons for Bayesian multilevel logistic regressions predicting likelihood of cue visits. Models only differed in their random effects. The fixed effects Effort, Predictive Validity, Trial, and their interactions, were kept the same for all models. The parameter of "Effort" is made up of −1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of Trial is made up of trial number / 270. The R syntax for each random effect structure is included. With the lowest WAIC value, Model 1 was chosen as the best fit. For comparison, the best fitting model with Predictive Validity as a continuous predictor had an WAIC of 14101.6

Parameter Estimates with Predictive Validity as a Categorical Variable

The parameter estimates for the best fitting model, with Predictive Validity as a categorical variable, can be seen in Table 11. Conditional Effects from this model are visibly closer to the raw data in Figure 22 (see Figure 25 for model conditional effects, and Figure 26 for conditional effects with effort back-transformed). The better fitting model using Predictive Validity as a categorical variable shows strong support for Hypothesis 1B. Though there is more variability in Experiment 2's data, it appears that 100% correct information drives cue visits more than any other predictive validity, even when it requires increased effort.

Table 11

Parameter Estimates for Bayesian Logistic Regression with Predictive Validity as a Categorical Variable Predicting Likelihood of Cue Visits for Experiment 2

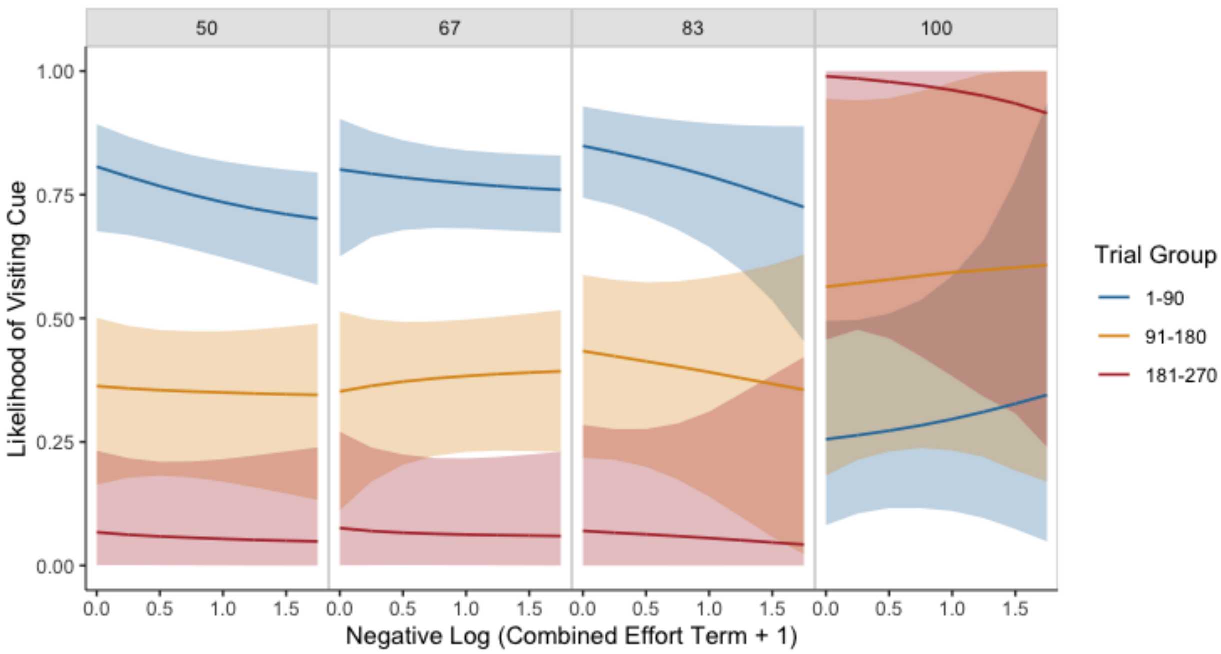
	B	L−95% CI	U−95% CI	*BF*
Intercept	1.15	−1.89	4.37	3.17
Effort	−5.56	−8.12	−2.93	>1000
Predictive Validity[67]	0.50	−1.27	2.30	2.47
Predictive Validity[83]	−1.05	−4.71	3.08	2.57
Predictive Validity[100]	24.00	8.98	39.83	>1000
Trial	−45.17	−55.88	−34.98	>1000
Block	3.45	1.83	5.09	>1000
Effort:Predictive Validity[67]	−0.42	−2.72	1.93	1.82
Effort:Predictive Validity[83]	1.93	−2.14	5.95	4.57
Effort:Predictive Validity[100]	1.19	−10.22	12.97	1.37
Effort:Trial	−5.71	−16.52	4.00	6.44
Predictive Validity[67]:Trial	3.87	−3.44	11.11	5.70
Predictive Validity[83]:Trial	−11.89	−27.15	1.07	26.30
Predictive Validity[100]:Trial	58.08	19.69	96.38	>1000
Effort:Block	2.39	0.89	3.95	959.00
Predictive Validity[67]:Block	−0.79	−1.89	0.29	12.27
Predictive Validity[83]:Block	3.31	1.23	5.51	>1000
Predictive Validity[100]:Block	−19.38	−25.96	−13.01	>1000
Trial:Block	2.94	1.57	4.32	>1000
Effort:Predictive Validity[67]:Trial	−9.49	−18.48	−0.95	69.18
Effort:Predictive Validity[83]:Trial	14.41	1.20	28.61	62.49
Effort:Predictive Validity[100]:Trial	18.77	−3.64	49.86	15.05
Effort:Predictive Validity[67]:Block	2.04	0.80	3.32	>1000
Effort:Predictive Validity[83]:Block	−2.73	−4.74	−0.71	213.29
Effort:Predictive Validity[100]:Block	1.22	−3.27	5.72	2.43
Effort:Trial:Block	−1.06	−2.49	0.41	13.14
Predictive Validity[67]:Trial:Block	0.31	−1.18	1.80	1.98
Predictive Validity[83]:Trial:Block	−1.89	−4.63	1.02	10.06
Predictive Validity[100]:Trial:Block	11.96	4.12	20.34	826.59
Effort:Predictive Validity[67]:Trial:Block	−0.84	−2.32	0.68	6.55
Effort:Predictive Validity[83]:Trial:Block	0.55	−2.24	3.27	1.89
Effort:Predictive Validity[100]:Trial:Block	−6.28	−12.37	−0.49	59.76

Note. Parameter estimates for the Bayesian multilevel logistic regression predicting likelihood of cue visits for Experiment 2. The parameter of "Effort" is made up of −1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity +

1) centered by its minimum. The labels of the Predictive Validity levels are written in their original scale (67, 83, 100). The categorical variable of Predictive Validity is dummy coded, with the minimum level 50 as the reference level. The parameter of Trial is made up of trial number / 270. This model included random intercept and slope effects of effort, predictive validity, trial, and all possible two-way interactions. Estimates are given in log odds space.

Figure 25

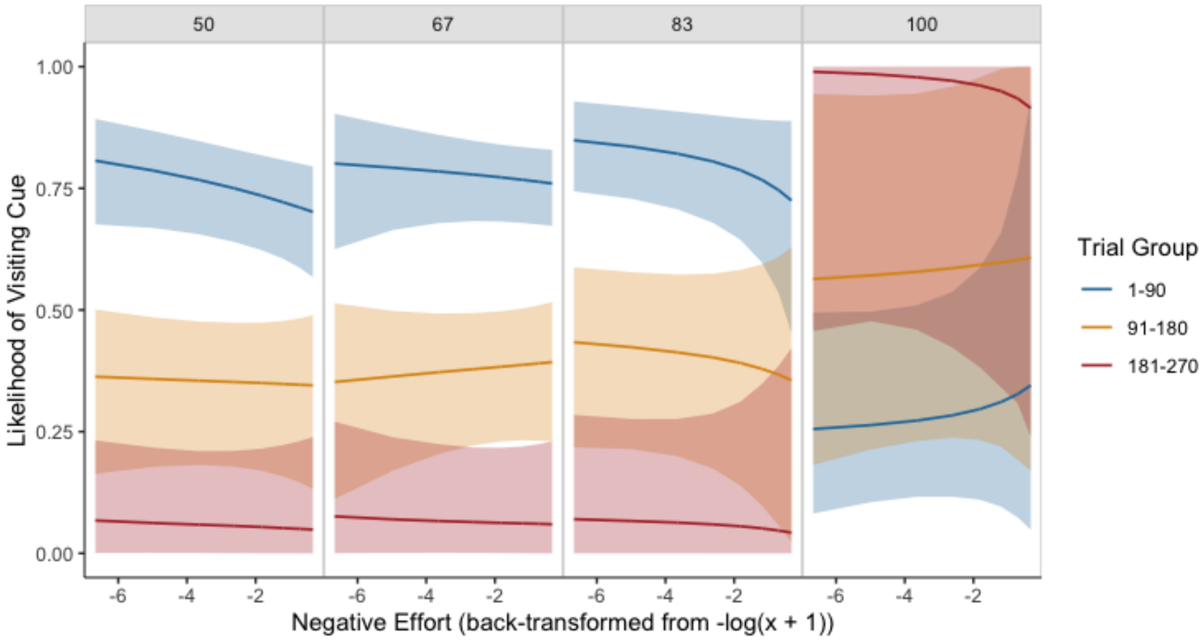
Conditional Effects of Logistic Regression for Experiment 2, with PV as a Categorical Predictor



Note. In this figure, the Effort variable appears as it did in the analyses, log transformed (log(x+1)), minimum-centered, and multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Credible Intervals.

Figure 26

Conditional Effects of Logistic Regression for Experiment 2, Categorical PV, Back Transformed



Note. In this figure, the Effort variable is back-transformed to its original scale but still multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Credible Intervals.

The best fitting model shows strong evidence for a negative effect of Effort (B = −5.56, 95% CI [−8.12,−2.93], BF > 1000), alongside a strong positive effect of predictive validity for the 100 cue (B = 24.00, 95% CI [8.98,39.83], BF > 1000), Similar to Experiment 1, these results support both Hypothesis 1A, the Effort Dominance Hypothesis, and Hypothesis 1B, the Predictive Validity Dominance Hypothesis, under different conditions, shaped by strong interactions such as Predictive Validity [100] and Trial (B = 58.08, 95% CI [19.69,96.38], BF > 1000) (for other significant interactions, see Table 11). A visualization of the likelihood of visiting the 50, 67, or 83 cue on trials where the 100 cue is not visited can be seen in Appendix L (1,929 instances, out of 8,370 total [31 participants × 270 trials]).

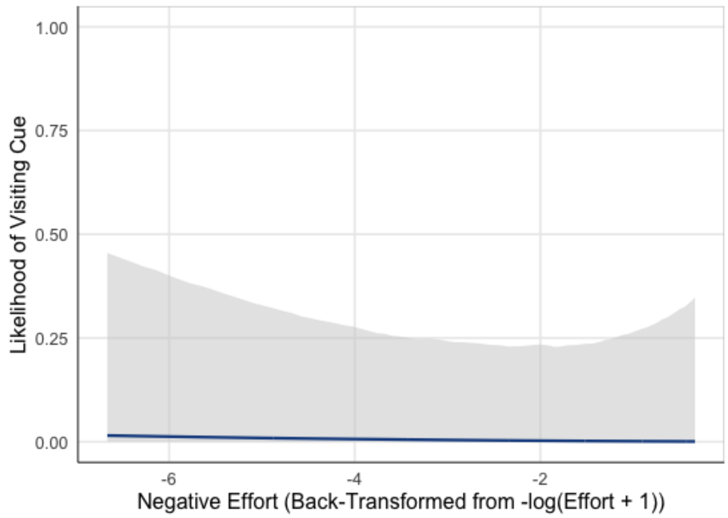
Visits to cues with predictive validities of 50, 67, and 83 appear to be slightly more strongly influenced by effort than the 100% cue but are also much more variable than cue visits in Experiment 1 (see large credible intervals; 50 PV: likelihood of cue visit at low effort of 0 for minimum centered −1(log(Effort + 1)), $M = 0.01$, 95% CI = [$8.48e^{-7}$, 0.27], higher than at high effort of 1.75, $M = 0.0005$, 95% CI = [$3.27e^{-11}$, 0.15]; $OR = 29.97$, 95% CI = [$8.81e^{-6}$, 6402]; 67 PV: likelihood of cue visit at low effort, $M = 0.02$, 95% CI = [$2.67e^{-7}$, 0.36] higher than at high effort, $M = 0.001$, 95% CI = [$3.64e^{-11}$, 0.41], $OR = 12.75$, 95% CI = [$1.34e^{-6}$, 3861]; 87 PV: likelihood of cue visit at low effort, $M = 0.05$, 95% CI [$9.64e^{-9}$, 0.94] higher than at high effort, $M = 0.005$, 95% CI = [$2.00e^{-13}$, 0.97], $OR = 10.21$, 95% CI = [$1.38e^{-9}$, 70734]; 100 PV: likelihood of cue visit at low effort, $M = 0.90$, 95% CI [$1.58e^{-6}$, 1.00] lower but closer in magnitude to that at high effort, $M = 0.99$, 95% CI = [$3.28e^{-07}$, 1.00], $OR = 0.03$, 95% CI = [0, 4417586]).

Predictive validity does not have a large effect on cue visits between the 50, 67, and 83 cues (for 50, with all other variables at their average value, M likelihood of cue visits = 0.002, 95% CI [$9.40e^{-9}$,0.122]; for 67, $M = 0.004$, 95% CI [$2.80e^{-9}$,0.238]; for 83, $M = 0.01$, 95% CI [$7.00e^{-11}$,0.869]) but the likelihood of visiting the 100 cue is much larger than all others ($M = 0.98$, 95% CI [$9.04e^{-6}$,1.00]; compared to 83, $OR = 3351$, 95% CI [$1.75e^{-15}$, $9.31e^{-9}$]). Compared to Experiment 1, these results have much wider credible intervals, indicating a large amount of variability in the results. Similar to Experiment 1, Experiment 2 provides support in the direction of both Hypotheses 1A and 1B under different conditions, but future study will aim to clarify the sources of variance when cue predictive validity values are masked from participants and must be learned through trial and error.

Similar to Experiment 1, Experiment 2 suggests a slight convex shape of effort's effect, supporting Hypothesis 3B (Largest Discounting Early). This conclusion is supported by the best fitting transforms for the effort variable, log transform (AIC = 26353.54) and cube root (AIC = 26342.04). A visualization of the conditional effects of the model with the effort predictor back transformed (but multiplied by −1 to aid in interpretation, with higher values meaning more effort), similarly to Experiment 1, shows no strong pattern of concave or convex shapes of the effort effect (see Figure 27). The conditional effect of effort when all other variables are set to their means similarly does not provide conclusive evidence as to the shape of effort's function (slope changing from −0.003 at the low effort value of −6 to −0.001 at the high effort value of −2). For the best fitting model with Predictive Validity as a continuous predictor, these slopes are only slightly different at −0.014 at the low effort value of −6 to −0.011 at the high effort value of −2, still suggesting a slight convex shape (see Figure 28). These results alongside those from Experiment 1 provide slight support for Hypothesis 3B, but the functional form of effort still requires further study.

Figure 27

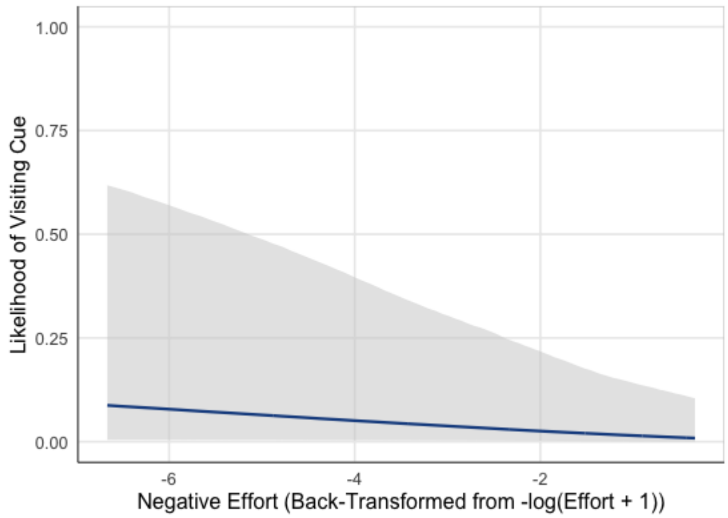
Conditional Effect of Effort for Logistic Regression with Categorical PV for Experiment 2



Note. Here, from the multilevel logistic regression predicting likelihood of cue visits for Experiment 2, with Predictive Validity as a categorical predictor, the conditional effect of Effort is shown when all other variables are at their means. The Effort variable is back-transformed to its original scale but still multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Error ribbon represents 95% Credible Interval.

Figure 28

Conditional Effect of Effort for Logistic Regression with Continuous PV, for Experiment 2



Note. Here, from the multilevel logistic regression predicting likelihood of cue visits for Experiment 2, the conditional effect of Effort is shown when all other variables are at their means. The Effort variable is back-transformed to its original scale but still multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Error ribbon represents 95% Credible Interval.

Hypotheses 2A and 2B were examined through comparison between Experiments 1 and 2, investigating whether or not effort has a greater effect when one has to learn cue predictive validities, compared to when predictive validity values are given to them.

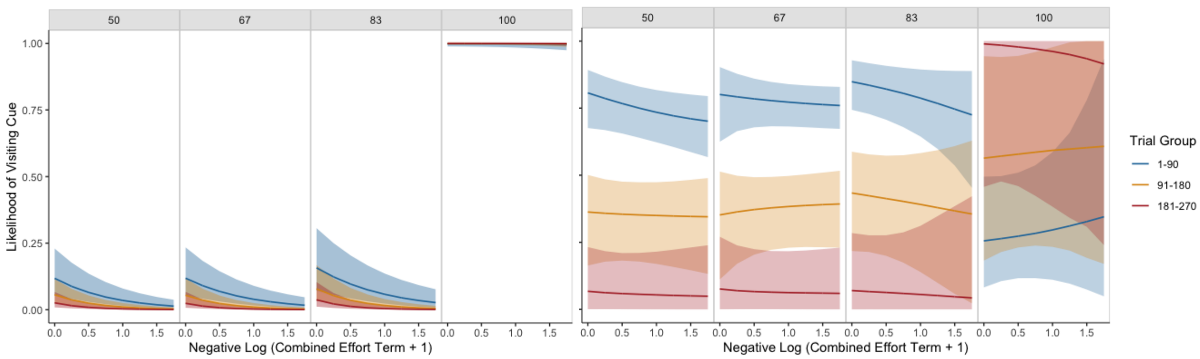
A comparison between the conditional effects for Experiments 1 and 2 shows support for Hypothesis 2A. This comparison is descriptive and not a formal statistical comparison, as Experiments 1 and 2 differ in their trial configurations, but this supports the need for a follow-up study directly looking at this effect.

These results suggest a slightly greater effect of effort in Experiment 2 compared to that found in Experiment 1 (Experiment 1 Effort $B = -1.13$, 95% CI [−1.65,−0.63], $BF > 1000$, compared to Experiment 2 Effort $B = -5.56$, 95% CI [−8.12,−2.93], $BF > 1000$)). Experiment 2 still shows a strong effect of Predictive Validity [100] ($B = 24.00$, 95% CI [8.98,39.83], $BF > 1000$), that is in fact larger than that of Experiment 1 ($B = 7.27$, 95% CI [5.27,9.86], $BF > 1000$), but the interactions and increased variability in Experiment 2 result in less strongly distinct estimated means across Predictive Validities compared to those from Experiment 1 (see the earlier reported means in this section).

The clearest takeaway from comparisons of Experiments 1 and 2 are that there is much more variability in responding when the predictive validity values are hidden from participants, and they must learn them through trial and error. Figure 29 shows the estimated conditional effects of the best fitting models from Experiments 1 and 2 side by side.

Figure 29

Conditional Effects of Logistic Regression, PV Categorical, for Experiment 1 vs Experiment 2



A. B.

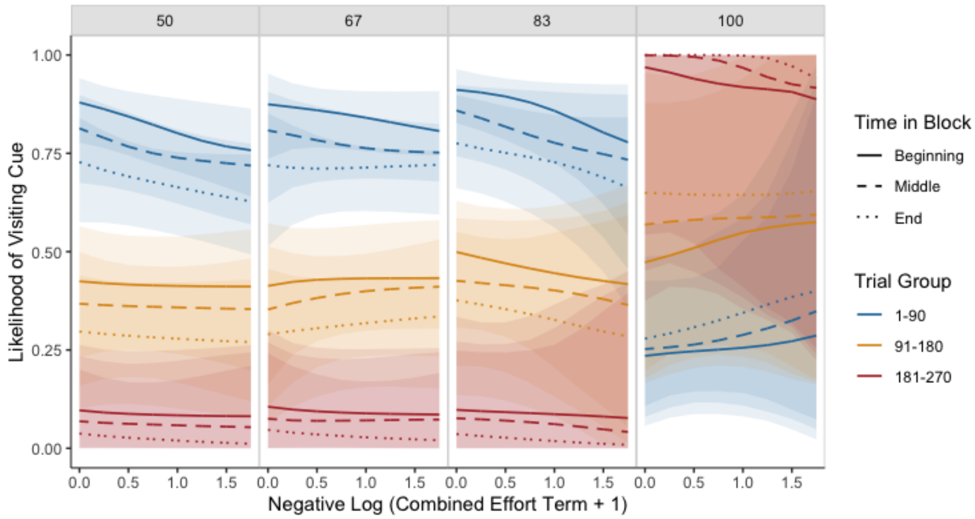
Note. In both A. (Experiment 1) and B. (Experiment 2), the Effort variable appears as it did in the analyses, log transformed (log(x+1)), minimum-centered, and multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Credible Intervals.

This difference in variability can additionally be seen in the comparisons of average cue visits for each Predictive Validity value and their credible intervals. For Experiment 1, PV 50, M = 0.01, 95% CI [0.004,0.03]; PV 67, M = 0.01, 95% CI [0.003,0.02]; PV 83, M = 0.02, 95% CI [0.004,0.04], and PV 100, M = 0.99, 95% CI [0.99,1.00]. These credible intervals are much smaller than those from Experiment 2, where PV 50, M = 0.002, 95% CI [$9.40e^{-9}$,0.122]; PV 67, M = 0.004, 95% CI [$2.80e^{-9}$,0.238]; PV 83, M = 0.01, 95% CI [$7.00e^{-11}$,0.869], PV 100, M = 0.98, 95% CI [$9.04e^{-6}$,1.00].

This variance is only partially explained by the presence of blocks in Experiment 2. Figure 30 visualizes the conditional effects plot for Experiment 2 with the lines for Trial (grouped by trials 1-90, 91-180, and 181-270) split into three lines each depicting the Beginning 15 trials, Middle 15 trials, and End 15 trials of each block (this visualization with effort back-transformed can be seen in Figure 31; for comparison, by-block figures for the best fitting model with Predictive Validity as a continuous predictor are in Appendix M). Although these lines show differences in cue visits over the course of a block, they are not as distinct as the differences from the 50%, 67%, and 83% cues in Experiment 1 to the 100% cue in that experiment (see Figures 14 and 15). The variability in responding in Experiment 2 can have different interpretations that will be expanded upon in the Discussion.

Figure 30

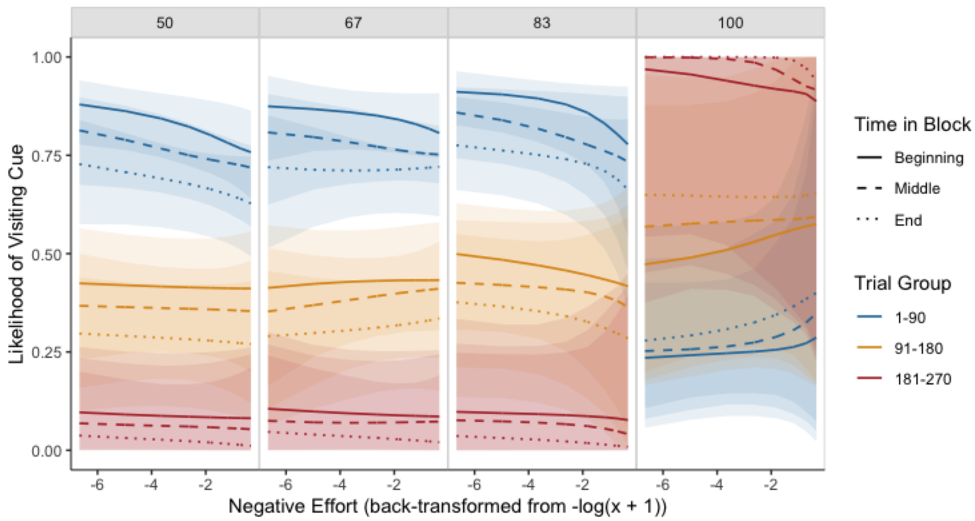
Conditional Effects of Logistic Regression for Experiment 2, Separated by Block



Note. In this figure, the Effort variable appears as it did in the analyses, log transformed (log(x+1)), minimum-centered, and multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Time in Block represents the Beginning 15 trials, Middle 15 trials, and End 15 trials of all blocks. Error ribbons represent 95% Credible Intervals.

Figure 31

Conditional Effects of Logistic Regression for Experiment 2, Separated by Block, Back-Transformed



Note. In this figure, the Effort variable is back-transformed to its original scale but still multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Time in Block represents the Beginning 15 trials, Middle 15 trials, and End 15 trials of all blocks. Error ribbons represent 95% Credible Intervals.

Chapter 4 - Discussion

The present study examined how effort and predictive validity shape the search for information in a probabilistic environment. Building from Brunswik's probabilistic functionalism (Brunswik, 1955), the search for information was conceptualized as the process of sampling proximal environmental cues under constraints of cost and uncertainty. Effort was introduced as a cost by manipulating the mouse cursor speed and cue distance in a Mouse-Contingent Bi-Resolution Display. Both experiments in this study investigated how effort costs modulate cue sampling, and how the predictive validity of proximal cues guides information search when validity information is either provided (Experiment 1) or learned through sampling and feedback from the environment (Experiment 2). Across both experiments, results demonstrated that effort limits cue exploration in favor of exploitation of the most predictive, 100% correct cue. The resilience of visits to the 100% cue even despite effort increases suggests a certainty effect for information cues.

Effort and Predictive Validity

The first set of hypotheses examined effort and predictive validity's unique roles in shaping cue visits. Both experiments provided evidence that increased effort decreases the likelihood of sampling non-perfectly-predictive cues. There is some evidence in support of Hypothesis 1A, the Effort Dominance Hypothesis. Visits to the 50%, 67%, and 83% cues were diminished by increased effort in both Experiments 1 and 2 (see results in Table 6 and Figure 15 for Experiment 1, and Table 11 and Figure 26 for Experiment 2). The findings of effort restricting visits to these three cues are consistent with Brunswik's view of perception as an ecologically constrained process (Brunswik, 1955), and consistent with existing theories about effort shaping behavior through a cost-benefit analysis (Bialaszek et al., 2017; Hull, 1943; Kool

& Botvinick, 2014, 2018; Kurniawan et al., 2010; Mitchell, 2017). In Experiment 2, visits to the 100% cue were also visibly impacted by a main effect of effort, albeit to a lesser degree (see comparison in Figure 29). Effort would occasionally shape cue visits towards less predictive cues, providing some evidence in support of Hypothesis 1B (see Appendices G and L for visualizations of cue visits in trials where the 100% cue was not visited), but overall, in both experiments the 100% cue was visited most often than any other cue.

In both experiments, utilization of the 100% cue remained highest despite the effort manipulation, and this utilization was highest in Experiment 1 where participants could see predictive validity values next to the cues (e.g., the 100% correct cue was accompanied by a poignant "100") instead of having to learn these values through trial and error. Highest utilization of the 100% cue in both experiments provides strong evidence in support of Hypothesis 1B, the Predictive Validity Dominance Hypothesis, where a need for certainty drives participants to seek out the most predictive information even at higher effort costs. This finding has similarities with the certainty effect found in probability discounting studies, where people overweight options they perceive as certain and underweight those they do not (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992; Webb & Young, 2015; Young et al., 2014). This finding also aligns with results from other Multiple Cue Probability Learning studies, where participants tend to use a maximizing strategy, relying on the most valid cue of the available options (Little & Lewandowsky, 2012; Simon, 1955, 1956; White & Koehler, 2007). The best fitting models from Experiments 1 and 2 were those that treated Predictive Validity as a categorical variable. The large effect sizes and Bayes factors for the 100% cue in both experiments, contrasted with the near-zero (Experiment 1) or highly variable (Experiment 2) estimates for the 50%, 67%, and 83% cues, demonstrate that participants treated 100% valid information as categorically distinct, a factor warranting future study.

The strong positive interaction between Predictive Validity and Trial in Experiment 2 also indicates that this certainty preference grows with learning. Participants gradually converged on the 100% cue as they accumulated evidence of its reliability. This result illustrates the self-reinforcing nature of valid information. Mackintosh (1975) and Le Pelley et al. (2011) discuss the cyclical relationship in which attention is drawn to predictive cues, and this attention increases our ability to learn about their predictive validity. The results of the 100% cue drawing more attention over the course of the study parallels this finding.

From these results, I can conclude that Hypothesis 1A (Effort Dominance) and 1B (Predictive Validity Dominance) were both supported under different conditions. Effort generally constrained exploration, while predictive validity drove exploitation of the 100% valid cue once its reliability was established. Across both experiments, positive interaction effects including Effort and Predictive Validity (see parameter estimates of best fitting model for Experiment 1 in Table 6 and for Experiment 2 in Table 11) suggest context-sensitive weighting consistent with the cost-benefit analyses described in the effort and decision-making literature (Kool & Botvinick, 2014, 2018). People appear to tolerate greater effort expenditure when the expected value justifies that cost. Within this experiment, that pattern is exhibited through participants selectively investing their effort towards perfectly predictive cues.

Learning vs. Known Predictive Validity

Differences in behavioral patterns observed between Experiments 1 and 2 showed that needing to learn cue validity information shaped the relative effects of effort and predictive validity. This pattern supported Hypothesis 2A (Effort Shaping Cue Learning) over Hypothesis

2B (Learning Similarly Despite Effort). Quantitatively, the main effect of Effort was larger in Experiment 2 ($B = -5.56$), than in Experiment 1 ($B = -1.13$), which indicates that effort deterred cue visits more strongly when cue validities had to be learned. This greater effect of effort was likely due to participants' inability to assess whether an effort investment would be justified before learning the predictive validities of the cues. Conversely, when predictive validities were explicitly provided in Experiment 1, participants could immediately weigh effort costs against known information value, leading to more efficient cue sampling.

When predictive validities were hidden in Experiment 2, both effort and predictive validity main effects remained strong but were accompanied by a substantial increase in between-subject variability. Several cognitive factors and individual differences likely contributed to this heterogeneity. First, participants may have differed in how quickly they inferred cue reliability. Some participants reached stable cue preferences early, while others continued sampling (see number of cue visits per trial in Figure 21), widening credible intervals for mean cue visits (see best-fitting model conditional effects in Figures 23 and 24). It is also possible that some participants may have shifted strategies over the course of the experiment. Second, subjective evaluation of effort likely differed across individuals. Kool and Botvinick (2018) noted that individuals have unique sensitivities to effort and reward, which has made it difficult to identify a single best-fitting effort discounting function. Although behaviors in Experiment 1 would also be subject to this source of variability, it is possible this variability compounded with the inability to immediately discern the value to be gained by expending effort in Experiment 2. Finally, the block structure of Experiment 2 required learning new cue information in each new block. The positive interaction effect of Effort and Block suggests some adaptation over time, where participants became less deterred by effort across blocks, perhaps reflecting habituation or improved motor efficiency. Conversely, the negative interaction effect between the 100% Predictive Validity cue and Block indicates that the effect of the perfectly predictive cue weakened slightly across blocks. This pattern could reflect decision fatigue, diminishing motivation as the experiment progressed, or satisfaction with viewing "good enough" cues (see the small positive interaction effect between the 83% Predictive Validity cue and Block). Individual reactions to the block structure could have contributed to the between-subject variability, beyond what was seen in the fully-randomized trials in Experiment 1.

The Shape of Effort

In the investigation of the shape of effort's functional form, both Experiments 1 and 2 supported a slightly convex shape for the effect of effort, closer to the functional form of delay discounting than some concave effort discounting functions. The best fitting log or cube-root transformations of the Effort variable (see Table 2 for a comparison of effort transformations for Experiment 1 and Table 7 for Experiment 2), and slight flattening of slopes at higher effort (see the conditional effect of Effort when all other variables were at their means, Figures 16 and 17 for Experiment 1 and Figures 27 and 28 for Experiment 2), indicate that overall participants showed steeper initial avoidance of effort followed by relative insensitivity to further increases. These results support Hypothesis 3B (Largest Discounting Early) over Hypothesis 3A (Largest Discounting Late). Historically, the shape of effort's discounting effect is less regularly replicated than that of delay discounting, however it is important to note that these mixed findings occur across many different effort operationalizations (Bialaszek et al., 2017; Klein-Flugge et al., 2015; Mitchell, 2017). In general, effort encompasses a wider range of possible manipulations than delay discounting. Consequently, differences in how effort is operationalized may yield functional forms of effort that are not entirely comparable across methodologies. It is also possible that different effort discounting studies are gathering piecewise views of a larger shape of effort's functional form. If a broader function of effort were to follow an S-shape like the weighting function in cumulative prospect theory (Tversky & Kahneman, 1992), overestimating small differences in effort could result in a convex decline, whereas underestimating larger differences in effort up until a point where estimates return to normal (say, for example, at the limits of physical ability) could result in a concave shape. Such a shape could only be identified through a study that examined a sufficiently wide range of "effort," potentially spanning multiple methodologies. Further study is certainly required to make any broader claims about the nature of effort's discounting function. For the time being, this study's findings regarding the shape of an effort effect adds to our understanding of what the shape of effort can look like, particularly in how effort shapes the search for cue information.

Overall, these findings support a view of information search as a process that continuously integrates and weighs expected information gain and effort cost to guide sampling in uncertain environments. These findings reflect the valuable (Dall et al., 2005) and attention-drawing nature of predictive information (Le Pelley et al., 2011; Mackintosh, 1975). Predictive information also resists effort discounting more than less valid information (see Figure 29 for comparison of Experiments 1 and 2), showing a certainty effect for information. These findings improve our understanding of the methods by which we acquire and utilize proximal cues in our environment.

Limitations

One limitation of this study is the sample size for Experiment 2. The initial power analysis from Experiment 1 was used to decide the sample size for Experiment 2, under the expectation that the effect of effort would be larger and likely easier to obtain in Experiment 2 with a similarly sized sample. The large variance in response strategies of participants in Experiment 2 was not anticipated but makes sense in retrospect given individual differences in learning rates and effort tolerances. Collecting more data for a design requiring learning could either reduce some of the overall variance or help to identify whether the variance comes from distinct patterns of behavior identifiable among subgroups of the subjects with a follow-up cluster analysis.

Another methodological limitation is that the measure of effort in this study, the mouse speed coefficient divided by distance of the cues from screen center, inevitably covaries with the time it takes to visit cues. Even through participants could move their mouse or trackpad quickly or with more force to increase their speed in a trial with a low mouse speed coefficient, this larger physical movement was an option for participants and not a forced response method. In Experiments 1 and 2, it was found that participants moved their mouse with only somewhat more force to compensate for the slower speeds, resulting in speeds that were faster than would be produced by the same movements multiplied by the mouse speed coefficient, but still slower than 100% speed (see average recorded mouse speed per trial for Experiment 1 in Appendix D, and for Experiment 2 in Appendix J). Future designs could hold trial duration constant or use an effort manipulation that is wholly distinguishable from time (e.g., forcibly squeezing hand grips for a set amount of time vs. holding hand grips lightly for that same amount of time, see Bialaszek et al., 2017; Mitchell, 1999).

Future Directions

Beyond these methodological refinements, future research could examine the effects of effort and predictive validity in more real-world contexts, such as medical decision-making, online information search, or education settings, in order to determine how these findings generalize to naturalistic information environments.

Future study could also examine predictive validities that approach, but do not reach, certainty. The present design contrasted a chance cue (50%) and two moderately reliable cues (67% and 83%) with a perfectly reliable cue (100%), revealing a categorical certainty effect for perfectly predictive information. Future studies could test higher but imperfect validities (e.g., 90, 95, 98%) to determine whether the apparent discontinuity reflects a true qualitative boundary or a steep but continuous nonlinear function (as would be suggested by the probability weighting function from cumulative prospect theory; Tversky & Kahneman, 1992). Such work would inform whether attention and effort are uniquely shaped by absolute certainty or whether there is a subjective evaluation of whether a cue's reliability has passed some decision threshold. Higher but imperfect validities would also more accurately reflect real-world high validity information, where cues are very rarely, if ever, 100% predictive.

Finally, future research could examine whether the value of the decision outcome influences how much effort people are willing to expend or how readily they accept less-valid cues. Individual trials or blocks could vary in reward for correct responses, or punishment for incorrect responses, to see whether higher stakes increase tolerance for effort costs or amplify the certainty effect, and if lower stakes weaken the certainty effect.

Closing Remarks

Overall, these two experiments provide converging evidence that attention is guided by weighting of the validity of information and the cost of effort required to obtain it. When cue validities are known, effort predictably shapes attention to information. When cue validities are initially unknown and must be learned, effort's deterrent effect intensifies, and individual differences in learning rate and effort sensitivity increase behavioral variability. In both contexts, the shape of effort's discounting effect appears to follow a convex form, where initial increases in effort sharply reduce information sampling earlier in the effort range, but additional effort increments produce diminishing marginal decreases later. These results extend Brunswik's probabilistic framework by specifying some of the mechanisms through which cue sampling is constrained in the environment. These results also demonstrate that information search is governed by principles similar to those shaping other forms of cost-sensitive choice. Finally, these experiments establish a flexible, replicable paradigm for studying information search under effort constraints, providing a foundation for future investigations into how cognitive and motivational factors shape what we learn about the probabilistic world around us.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.

Anderson, J. R. (2013). *The adaptive character of thought*. Psychology Press. https://www.taylorfrancis.com/books/mono/10.4324/9780203771730/adaptive-character-thought-john-anderson

Anwyl-Irvine, A., Armstrong, T., & Dalmaijer, E. (2021). *MouseView. js: Reliable and valid attention tracking in web-based experiments using a cursor-directed aperture.*

Baddeley, A. D. (2007). *Working memory, thought, and action* (Vol. 8). Oxford University Press. https://books.google.com/books?hl=en&lr=&id=DRIeAAAAQBAJ&oi=fnd&pg=PT14&dq=(Baddeley,+2007%3B+&ots=Xtz4KVoJ79&sig=2zx_zqYE3kdGgC4O-gdG67jx1cY

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66–80.

Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends in Neurosciences, 34*(4), 210–224. https://doi.org/10.1016/j.tins.2011.02.003

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. *convergence, 12*(1), 2.

Bautista, L. M., Tinbergen, J., & Kacelnik, A. (2001). To walk or to fly? How birds choose among foraging modes. *Proceedings of the National Academy of Sciences, 98*(3), 1089–1094. https://doi.org/10.1073/pnas.98.3.1089

Beedie, C. J., & Lane, A. M. (2012). The Role of Glucose in Self-Control: Another Look at the Evidence and an Alternative Conceptualization. *Personality and Social Psychology Review, 16*(2), 143–153. https://doi.org/10.1177/1088868311419817

Białaszek, W., Marcowski, P., & Ostaszewski, P. (2017). Physical and cognitive effort discounting across different reward magnitudes: Tests of discounting models. *PloS One, 12*(7), e0182353.

Blackwell, A. F., Jansen, A. R., & Marriott, K. (2000). Restricted focus viewer: A tool for tracking visual attention. *International Conference on Theory and Application of Diagrams,* 162–177.

Botvinick, M. M., Huffstetler, S., & McGuire, J. T. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 16–27. https://doi.org/10.3758/CABN.9.1.16

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods, 23*(3), 389.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*(3), 193.

Bruya, B. (2010). *Effortless attention: A new perspective in the cognitive science of attention and action.* MIT Press. https://books.google.com/books?hl=en&lr=&id=QicTDgAAQBAJ&oi=fnd&pg=PR7&dq=Bruya+Effortless+Attention:+&ots=kQTkBpusnV&sig=mu2LyWtl4M8eEeupFWk1N6T9v0U

Bruya, B., & Tang, Y.-Y. (2018). Is Attention Really Effort? Revisiting Daniel Kahneman's Influential 1973 Book Attention and Effort. *Frontiers in Psychology, 9*, 1133. https://doi.org/10.3389/fpsyg.2018.01133

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software, 80*, 1-28.

Carter, E. C., Kofler, L. M., Forster, D. E., & McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology: General, 144*(4), 796.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Cowan, N. (1999). An embedded-processes model of working memory. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control, 20*(506), 1013–1019.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review, 24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281.

Croxson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E., & Rushworth, M. F. (2009). Effort-based cost–benefit valuation and the human brain. *Journal of Neuroscience, 29*(14), 4531–4541.

Dall, S. R., Giraldeau, L.-A., Olsson, O., McNamara, J. M., & Stephens, D. W. (2005). Information and its use by animals in evolutionary ecology. *Trends in Ecology & Evolution, 20*(4), 187–193.

Doherty, M. E., & Kurz, E. M. (1996). Social Judgement Theory. *Thinking & Reasoning, 2*(2–3), 109–140. https://doi.org/10.1080/135467896394474

Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science, 11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Fechner, G. T. (1948). *Elements of psychophysics, 1860*. https://psycnet.apa.org/record/2006-10213-026

Gray, J. C., & MacKillop, J. (2015). Impulsive delayed reward discounting as a genetically-influenced target for drug abuse prevention: A critical evaluation. *Frontiers in Psychology, 6,* 1104.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin, 130*(5), 769.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136*(4), 495.

Harris, J. K. (2021). Primer on binary logistic regression. *Family Medicine and Community Health, 9*(Suppl 1). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8710907/

Hartmann, M. N., Hager, O. M., Tobler, P. N., & Kaiser, S. (2013). Parabolic discounting of monetary rewards by physical effort. *Behavioural Processes, 100*, 192–196.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16*(4), 219–222.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4*(3), 267.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in cognitive sciences*, *13*(12), 517-523.

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., & Couzin, I. D. (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences, 19*(1), 46–54.

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory.* https://psycnet.apa.org/record/1944-00022-000

Inzlicht, M., & Gutsell, J. N. (2007). Running on Empty: Neural Signals for Self-Control Failure. *Psychological Science, 18*(11), 933–937. https://doi.org/10.1111/j.1467-9280.2007.02004.x

Inzlicht, M., & Schmeichel, B. J. (2016). Beyond limited resources: Self-control failure as the product of shifting priorities. *Handbook of Self-Regulation, 3*, 165–181.

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in cognitive sciences, 22*(4), 337-349.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259.

Jiang, M., Huang, S., Duan, J., & Zhao, Q. (2015). Salicon: Saliency in context. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 1072–1080.

Jones, M. N., & Mewhort, D. J. K. (2004). Tracking attention with the focus-window technique: The information filter must be calibrated. *Behavior Research Methods, Instruments, & Computers, 36*(2), 270–276.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*(4), 441–480.

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. In L. C. MacLean & W. T. Ziemba, *World Scientific Handbook in Financial Economics Series* (Vol. 4, pp. 99–127). WORLD SCIENTIFIC. https://doi.org/10.1142/9789814417358_0006

Kim, N. W., Bylinskii, Z., Borkin, M. A., Gajos, K. Z., Oliva, A., Durand, F., & Pfister, H. (2017). BubbleView: An interface for crowdsourcing image importance maps and tracking visual attention. *ACM Transactions on Computer-Human Interaction (TOCHI), 24*(5), 1–40.

Klein-Flügge, M. C., Kennerley, S. W., Saraiva, A. C., Penny, W. D., & Bestmann, S. (2015). Behavioral modeling of human choices reveals dissociable effects of physical effort and temporal delay on reward devaluation. *PLoS Computational Biology, 11*(3), e1004116.

Kool, W., & Botvinick, M. (2013). The intrinsic cost of cognitive control. Behavioral and Brain Sciences, 36(6), 697–698.

Kool, W., & Botvinick, M. (2014). *A labor/leisure tradeoff in cognitive control.* https://psycnet.apa.org/journals/mot/1/S/3/

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour, 2(*12), 899–908.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*(10), 1292–1298.

Kruschke, J. K. (2003). Attention in Learning. *Current Directions in Psychological Science, 12*(5), 171–175. https://doi.org/10.1111/1467-8721.01254

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Kurniawan, I. T., Seymour, B., Talmi, D., Yoshida, W., Chater, N., & Dolan, R. J. (2010). Choosing to Make an Effort: The Role of Striatum in Signaling Physical Effort of a Chosen Action. *Journal of Neurophysiology, 104*(1), 313–321. https://doi.org/10.1152/jn.00027.2010

Kurzban, R. (2010). Does the Brain Consume Additional Glucose during Self-Control Tasks? *Evolutionary Psychology, 8*(2), 244–259. https://doi.org/10.1177/147470491000800208

Kurzban, R. (2016). The sense of effort. *Current Opinion in Psychology, 7*, 67–70.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences, 36*(6), 661–679.

Lakens, D. (2022). Sample size justification. *Collabra: Psychology, 8*(1), 33267.

Le Pelley, M. E., Beesley, T., & Griffiths, O. (2011). Overt attention and predictiveness in human contingency learning. *Journal of Experimental Psychology: Animal Behavior Processes, 37*(2), 220.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Emmeans: estimated marginal means, aka least-squares means (Version 1.3. 4). *Emmeans Estim Marg Means Aka Least-Sq Means, 4.*

Lewis, M. (1964). Effect of effort on value: An exploratory study of children. *Child Development,* 1337–1342.

Little, D. R., & Lewandowsky, S. (2012). Multiple-cue probability learning. *Encyclopedia of the Sciences of Learning,* 2386–2388.

Loewenstein, G. F. (1988). Frames of Mind in Intertemporal Choice. *Management Science, 34*(2), 200–214. https://doi.org/10.1287/mnsc.34.2.200

MacArthur, R. H., & Pianka, E. R. (1966). On Optimal Use of a Patchy Environment. *The American Naturalist, 100*(916), 603–609. https://doi.org/10.1086/282454

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*(4), 276.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., Hausmann, D., Fiedler, K., & Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision, 2*(3), 191–215. https://doi.org/10.1037/dec0000033

Mitchell, S. H. (1999). Measures of impulsivity in cigarette smokers and non-smokers. *Psychopharmacology, 146*, 455–464.

Mitchell, S. H. (2017). Devaluation of Outcomes Due to Their Cost: Extending Discounting Models Beyond Delay. In J. R. Stevens (Ed.), *Impulsivity* (Vol. 64, pp. 145–161). Springer International Publishing. https://doi.org/10.1007/978-3-319-51721-6_5

Mormann, M., & Russo, J. E. (2021). Does attention increase the value of choice alternatives? *Trends in Cognitive Sciences, 25*(4), 305–315.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 28*(3), 411.

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation, 51*, 45–100.

Oberauer, K. (2019). Working memory and attention–A conceptual analysis and review. *Journal of Cognition, 2*(1). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6688548/

Orquin, J. L., & Kurzban, R. (2016). A meta-analysis of blood glucose effects on human decision making. *Psychological Bulletin, 142*(5), 546.

Orquin, J. L., Lahm, E. S., & Stojić, H. (2021). The visual environment and attention in decision making. *Psychological Bulletin, 147*(6), 597.

Paas, F., Van Gog, T., & Sweller, J. (2010). Cognitive Load Theory: New Conceptualizations, Specifications, and Integrated Research Perspectives. *Educational Psychology Review, 22*(2), 115–121. https://doi.org/10.1007/s10648-010-9133-8

Payne, J. W. (1993). *The adaptive decision maker*. Cambridge University Press. https://books.google.com/books?hl=en&lr=&id=QzXFqwrPLXkC&oi=fnd&pg=PR11&dq=payne,+bettman,+and+johnson+1993&ots=14NP2bDCil&sig=lvfUiN1nBzE7-LA3z5yQd0v1pzE

Payne, K. (2023). *Online mouse tracking as a measure of attention in videos, using a mouse-contingent bi-resolution display* [Masters Thesis]. https://krex.k-state.edu/handle/2097/43472

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods, 54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Rachlin, H., Raineri, A., & Cross, D. (1991). SUBJECTIVE PROBABILITY AND DELAY. *Journal of the Experimental Analysis of Behavior, 55*(2), 233–244. https://doi.org/10.1901/jeab.1991.55-233

Rodd, J. M. (2024). Moving experimental psychology online: How to obtain high quality data when we can't see our participants. *Journal of Memory and Language, 134*, 104472.

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rothstein, H. G. (1986). The effects of time pressure on judgment in multiple cue probability learning. *Organizational Behavior and Human Decision Processes, 37*(1), 83–92.

Schouppe, N., Ridderinkhof, K. R., Verguts, T., & Notebaert, W. (2014). Context-specific control and context selection in conflict tasks. *Acta Psychologica, 146*, 63–66.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 461–464.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134*(2), 207.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience, 40*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics,* 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129–138.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics, 50*(3), 665–690.

Stoltzfus, J. C. (2011). Logistic Regression: A Brief Primer. *Academic Emergency Medicine, 18*(10), 1099–1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x

Sugiwaka, H., & Okouchi, H. (2004). Reformative self-control and discounting of reward value by delay or effort1. *Japanese Psychological Research, 46*(1), 1–9. https://doi.org/10.1111/j.1468-5884.2004.00231.x

Summerside, E. M., & Ahmed, A. A. (2021). Using metabolic energy to quantify the subjective value of physical effort. *Journal of The Royal Society Interface, 18*(180), 20210387. https://doi.org/10.1098/rsif.2021.0387

Tolman, E. C. (1932). *Purposive behavior in animals and men*. Appleton-Century. https://books.google.com/books?hl=en&lr=&id=U_Y63YLp8V0C&oi=fnd&pg=PA3&dq=purposive+behavior+in+animals+and+men&ots=xlPFDctzMQ&sig=rwSjyi_PmcOt2J8AIU_8Gsv3gdA

Tooby, J., & Cosmides, L. (2008). *The evolutionary psychology of the emotions and their relationship to internal regulatory variables*. https://psycnet.apa.org/record/2008-07784-008

Treadway, M. T., Buckholtz, J. W., Schwartzman, A. N., Lambert, W. E., & Zald, D. H. (2009). Worth the 'EEfRT'? The effort expenditure for rewards task as an objective measure of motivation and anhedonia. *PloS One, 4*(8), e6598.

Tversky, A., & Kahneman, D. (1974). Judgement under Uncertainty: Heuristics and Biases. *Science, 185*, 1124–1131.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323. https://doi.org/10.1007/BF00122574

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing, 27*(5), 1413-1432.

von Kortzfleisch, V. T., & Richter, S. H. (2024). Systematic heterogenization revisited: Increasing variation in animal experiments to improve reproducibility? *Journal of Neuroscience Methods, 401*, 109992.

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods, 17*(2), 228.

Webb, T. L., & Young, M. E. (2015). Waiting when both certainty and magnitude are increasing: Certainty overshadows magnitude. *Journal of Behavioral Decision Making, 28*(3), 294–307.

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience, 15*(2), 395–415. https://doi.org/10.3758/s13415-015-0334-y

White, C. M., & Koehler, D. J. (2007). Choice strategies in multiple-cue probability learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 757.

Wickens, C. D. (2014). Effort in Human Factors Performance and Decision Making. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 56*(8), 1329–1336. https://doi.org/10.1177/0018720814558419

Wickens, C. D. (2015). *Noticing events in the visual workplace: The SEEV and NSEEV models.* https://psycnet.apa.org/record/2015-10765-008

Wileyto, E. P., Audrain-Mcgovern, J., Epstein, L. H., & Lerman, C. (2004). Using logistic regression to estimate delay-discounting functions. *Behavior Research Methods, Instruments, & Computers, 36*(1), 41–51. https://doi.org/10.3758/BF03195548

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General, 143*(6), 2074.

Yarbus, A. L. (1967). Eye Movements During Perception of Complex Objects. In A. L. Yarbus, *Eye Movements and Vision* (pp. 171–211). Springer US. https://doi.org/10.1007/978-1-4899-5379-7_8

Young, M. E. (2018). Discounting: A practical guide to multilevel analysis of choice data. *Journal of the Experimental Analysis of Behavior, 109*(2), 293–312. https://doi.org/10.1002/jeab.316

Young, M. E., Webb, T. L., Rung, J. M., & McCoy, A. W. (2014). Outcome probability versus magnitude: When waiting benefits one at the cost of the other. *PloS One, 9*(6), e98996.

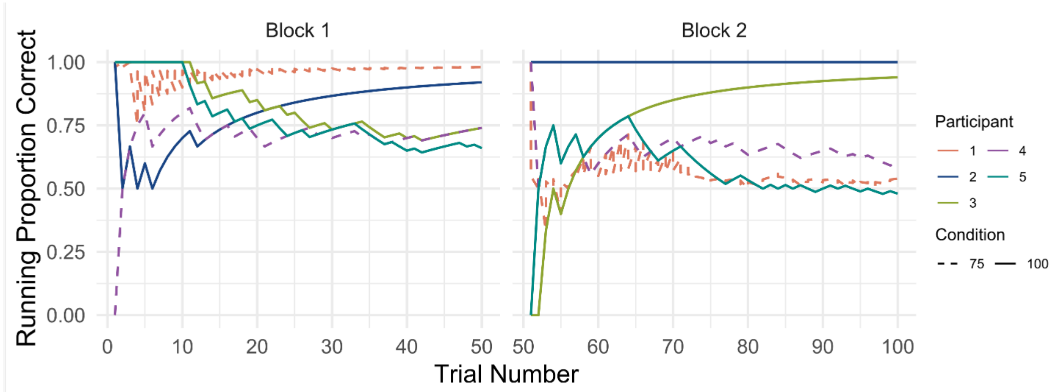
Appendix A - Early Pilots

Two early pilot experiments were run, largely as a proof-of-concept and a test of the methodology among naïve participants. These pilots were not run with sufficient participants for statistical analysis, rather they were run to determine what constituted low or high levels of effort, and the approximate number of trials it takes participants to identify cue predictive validities.

Pilot 1 consisted of five participants. Participants were presented with one practice trial to get brief experience moving the high-resolution window around the blurred display to visit cues. Participants then experienced 50 trials of low effort (100% speed), where one cue had 100% predictive validity (maximally informative), and the three other cues had 50% predictive validity (not informative). Two of the five participants identified the maximally predictive cue and were able to get 100% correct answers by trial 15 (see Figure 32 for the running proportion of correct responses for each participant). The other three participants did not. In Block 2 of the first pilot, the global effort was increased to high global effort (20% speed). Three participants (randomly assigned) got the same 100-50-50-50 distribution of cue predictive validities, and two participants got 75-50-50-50. Of those participants who still had a 100% cue, two of the three learned the maximally predictive cue and got all correct responses past trial 10. Of the two participants whose most valid cue was only 75%, all participants trended towards chance responses (50% correct). This pilot spurred the addition of more detailed instruction at the beginning of the experiment, plus the instantiation of a train-to-criterion practice block instead of one practice trial.

Figure 32

Running Proportion of Correct Responses from Participants in Pilot 1



Note. The line representing each participant's running proportion of correct responses will deflect downwards when that participant makes an incorrect response.

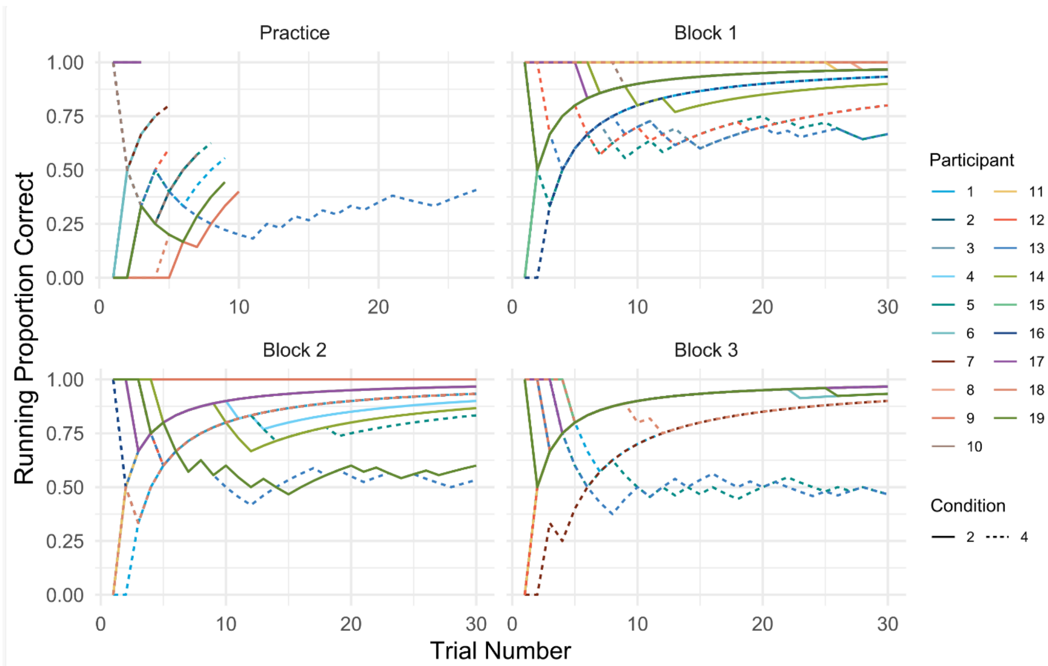
The second pilot had 19 participants, with 10 of them (randomly assigned) given only two cue options (left and right) throughout the experiment, and the remaining 9 participants getting the standard four cues. Each block in this pilot featured one 100% predictive validity cue, while the others only had 50% predictive validity. Participants experienced a practice block where they continued until they answered three trials correctly in a row. Following this practice block, participants experienced one 30-trial block with low global effort (100% speed) and the 100% cue being near the center, then a second 30-trial block with low global effort and the 100% cue being far from the center, followed by one final 30-trial block with high global effort (20% speed) and the 100% cue being far from the center.

In this pilot, most participants learned and solely visited the 100% correct cue, even in the effortful condition. Figure 33 shows the participants' running proportion of correct responses during each block, and Figure 34 shows the number of cues visited by each participant per trial. In Block 1, 15/19 participants identified and solely relied on the 100% cue by trial 13. In Block

2, 16/19 participants identified and solely relied on the 100% cue by trial 13, and in Block 3 (the block with high global effort), 17/19 participants identified and solely relied on the 100% cue by trial 13. Participants with four cue options did appear to take longer to identify the 100% cue. Pilot 2 shows better performance following the implementation of more detailed instructions and a train-to-criterion pilot block.

Figure 33

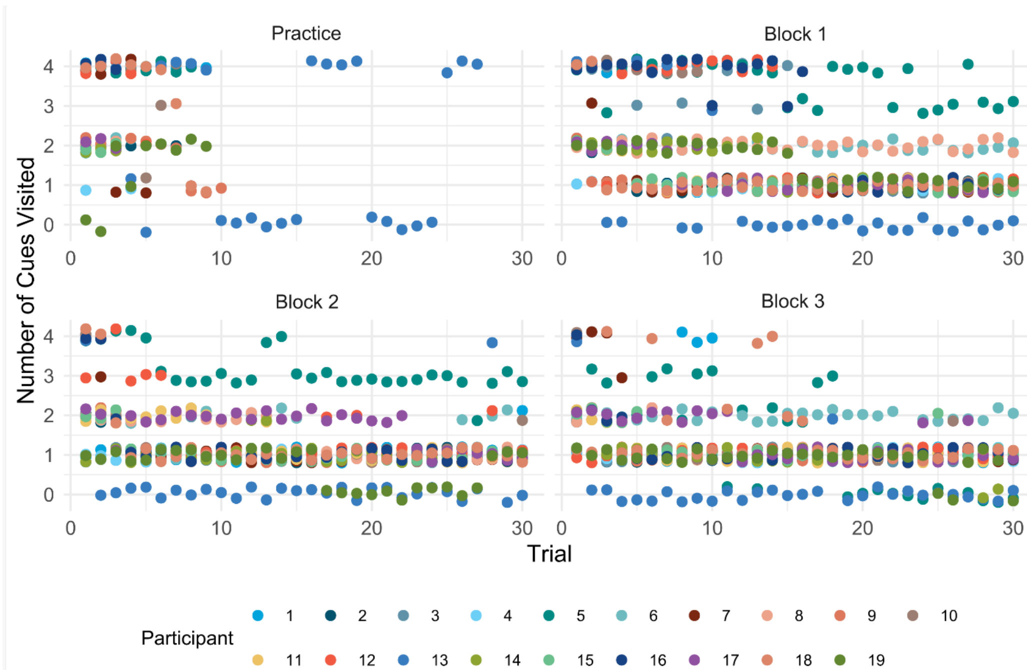
Running Proportion of Correct Responses from Participants in Pilot 2



Note. The line representing each participant's running proportion of correct responses will deflect downwards when that participant makes an incorrect response.

Figure 34

Number of Cues Visited by Participants in Pilot 2

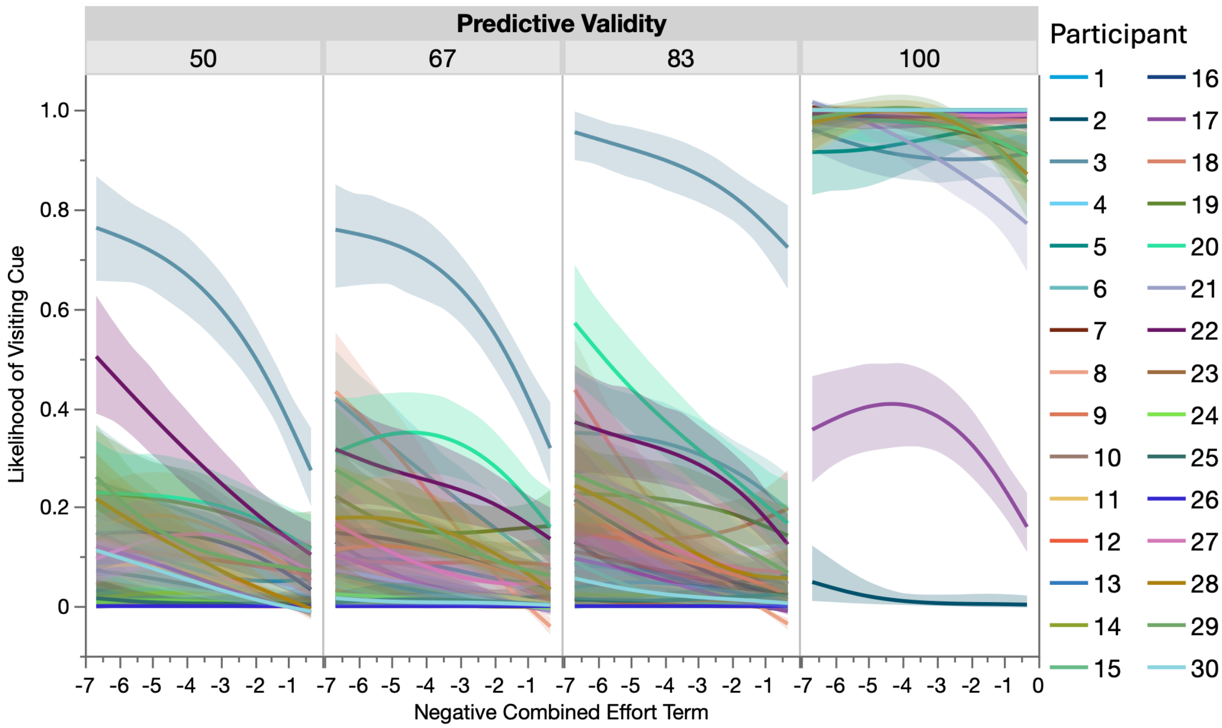


Note. Number of cues visited by each participant in each block of Pilot 2, by trial. By the end of the trials in a block, participants gravitate towards viewing only one cue.

Appendix B - Additional Raw Data Visualizations for Experiment 1

Figure 35

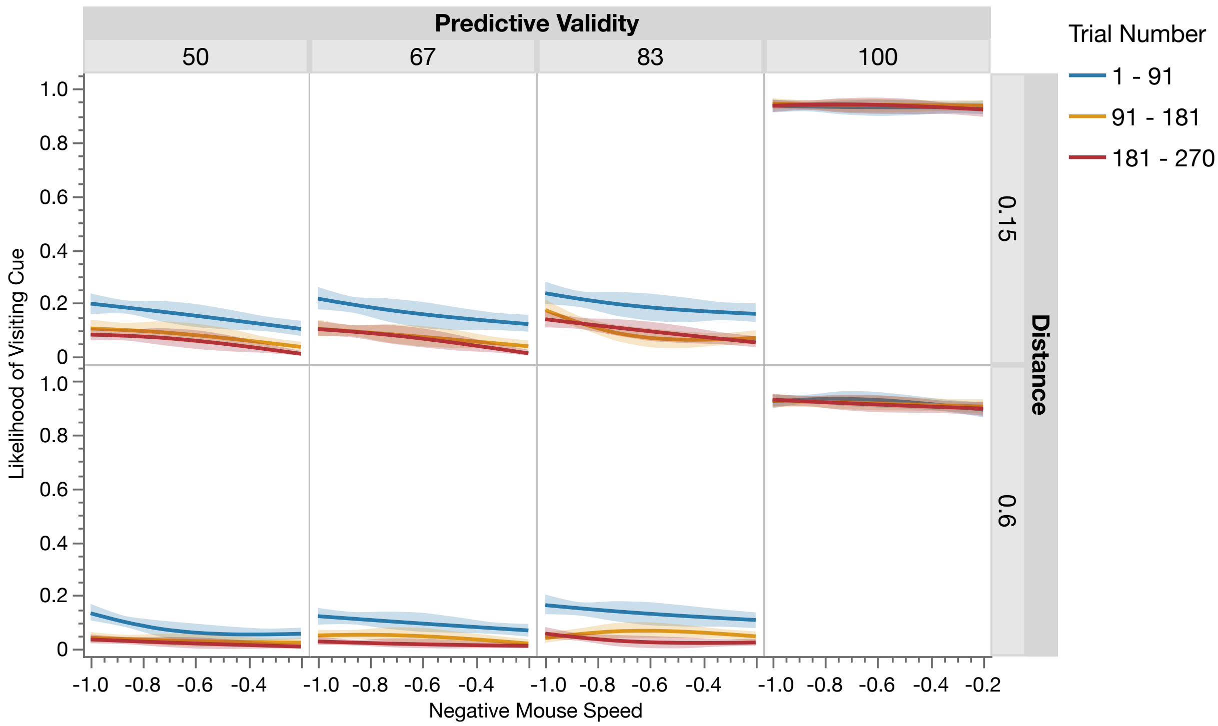
Raw Data for Experiment 1 by Effort, Predictive Validity, and Participant



Note. As all original responses are 0 (cue not visited on a trial) or 1 (cue visited on a trial), raw data is visualized here with a cubic spline fit (λ = 22 to reduce excessive curvature and predictions above 1 and below 0; JMP Student Edition 18.2.1). Any predicted values above 1 and below 0 reflect the spline's interpolation rather than possible likelihoods in the data. Error ribbons represent JMP's bootstrapped confidence region for the spline fits.

Figure 36

Raw Data for Experiment 1 by Mouse Speed, Distance, Predictive Validity, and Trial.

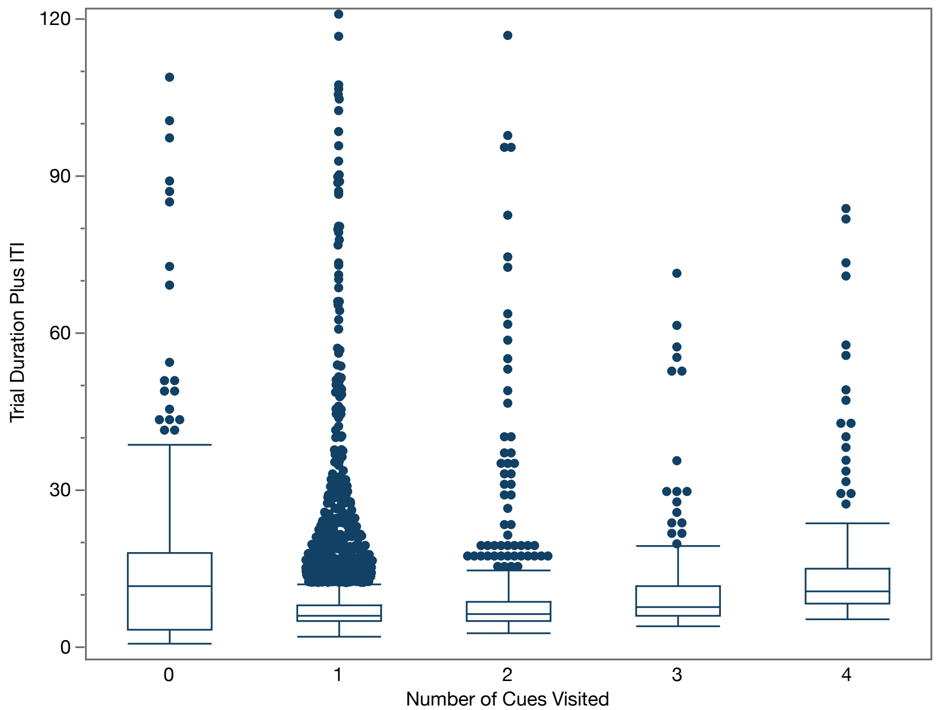


Note. In this figure the combined effort term is split into its component parts of mouse speed coefficient (multiplied by −1 such that lower, more effortful speeds reflect higher values) and distance (in proportion of screen height). This visualization shows no large discrepancies in responses that would caution against the usage of a combined effort term. As all original responses are 0 (cue not visited on a trial) or 1 (cue visited on a trial), raw data is visualized here with a cubic spline fit ($\lambda = 0$, i.e., no smoothing; JMP Student Edition 18.2.1). Error ribbons represent JMP's bootstrapped confidence region for the spline fits.

Appendix C - Trial Duration Visualizations for Experiment 1

Figure 37

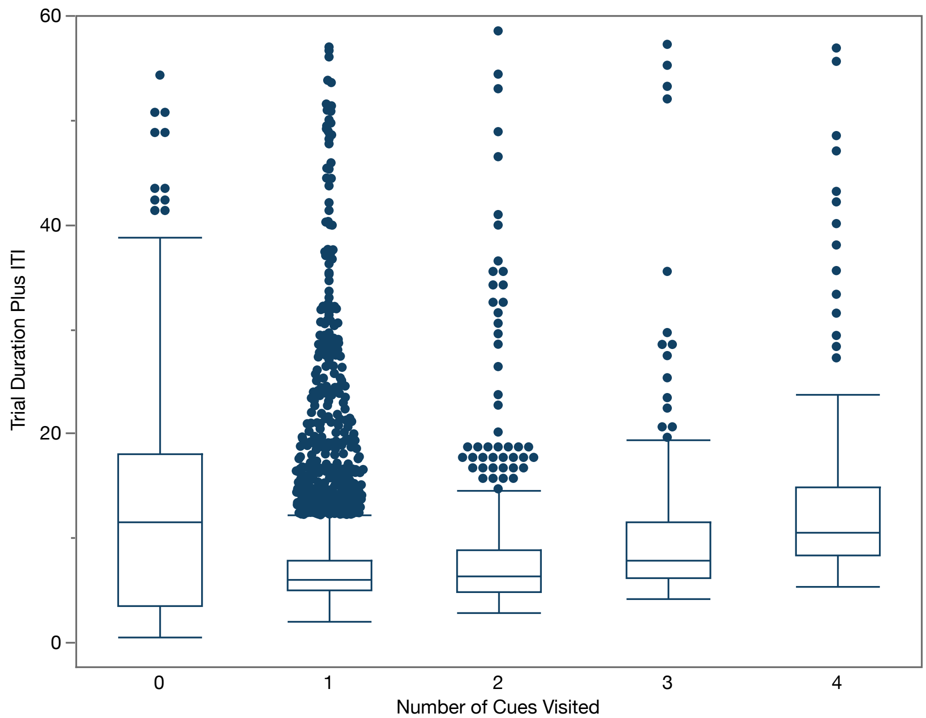
Trial Duration (Including ITI) by Number of Cues Visited for Experiment 1.



Note. This figure shows all 8100 trials experienced by participants (30 participants with 270 trials each), minus 37 trials with durations longer than 120 s.

Figure 38

Trial Duration (Including ITI) by Number of Cues Visited for Experiment 1, Zoomed In



Note. This figure shows all 8100 trials experienced by participants (30 participants with 270 trials each), minus 92 trials with durations longer than 60 s.

Table 12

Descriptive Statistics for Trial Duration (Including ITI)

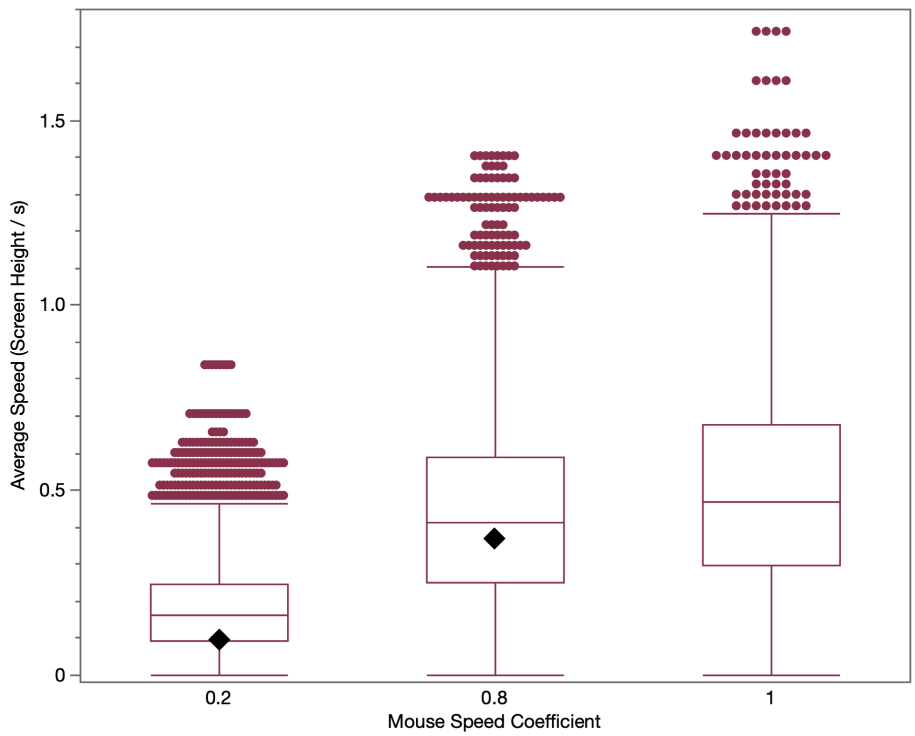
by Number of Cues Visited for Experiment 1

Number of Cues Visited	Mdn	IQR
0	11.55	[3.49, 17.97]
1	6.06	[4.95, 7.83]
2	6.39	[4.93, 8.82]
3	7.82	[6.17, 11.70]
4	10.56	[8.37, 14.91]

Appendix D - Average Speed by Mouse Setting, Experiment 1

Figure 39

Trial Average Speed by Mouse Speed Coefficient Setting for Experiment 1

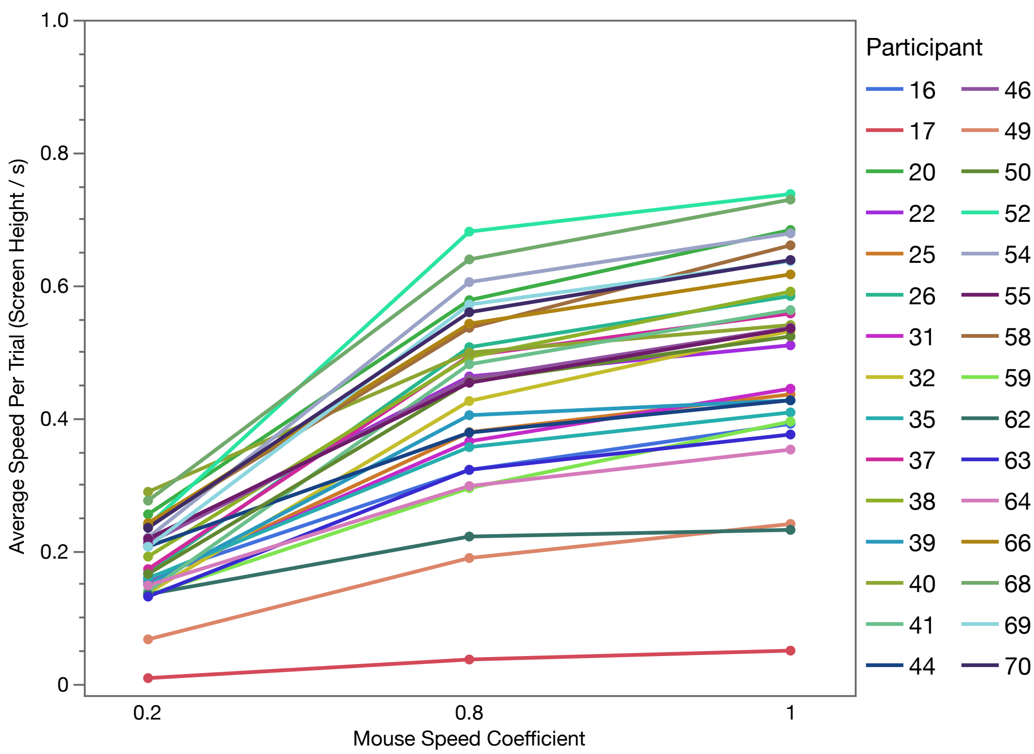


Note. This figure shows the average speed in screen height units per second for the 8100 trials experienced by participants (30 participants with 270 trials each). When the mouse speed was unchanged from the participants' normal speed, participants averaged a median of 0.47 screen height / s (*IQR* = [0.30, 0.68]). Diamonds on the plots for the 0.2 and 0.8 speed coefficients represent the anticipated outcome speeds relative to this 100% speed outcome (80% × 0.47 = 0.38; 20% × 0.49 = 0.09). When the mouse speed was set to 80% of its original speed, participants moved at a median speed of 0.41 screen height / s (*IQR* = [0.25 , 0.59]; 87% of the median for 100% speed). This faster outcome speed is possible by moving the mouse or trackpad physically more in a shorter amount of time than was done at the 100% speed setting.

Participants appeared to make even larger physical movements to compensate for the slowest mouse speed of 20%, with participants moving a median of 0.16 screen height / s (*IQR* = [0.09 , 0.25]; 34% of the median for 100% speed).

Figure 40

Trial Average Speed by Mouse Speed Coefficient Setting for Experiment 1, by Participant



Note. This figure shows the average speed in screen height units per second for each participant (90 trials per speed coefficient per participant).

Appendix E - Frequentist Models, Continuous PV, Experiment 1

Table 13

Model Comparisons for Logistic Regressions Predicting Likelihood of Cue Visits for Experiment 1

Model	Random Effects Syntax	AIC
1. Intercept Only	(1 | Participant)	19164.27
2. Only Main Effects	(1 + Effort+ Predictive Validity + Trial | Participant)	15625.01
3. Main Effects and Two-Way Interactions	(1 + Effort + Predictive Validity + Trial + Effort : Predictive Validity + Effort : Trial + Predictive Validity : Trial | Participant)	14504.86

Note. Model comparisons for multilevel logistic regressions predicting likelihood of cue visits. Models only differed in their random effects. The fixed effects Effort, Predictive Validity, Trial, and their interactions, were kept the same for all models. The parameter of "Effort" is made up of -1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of Trial is made up of trial number / 270. The R syntax for each random effect structure is included. A model with a three-way interaction in the random effects was not included as the model did not converge with bobyqa, nloptr, or Nelder-Mead optimizers. With the lowest AIC value, Model 3 was chosen as the best fit.

Table 14

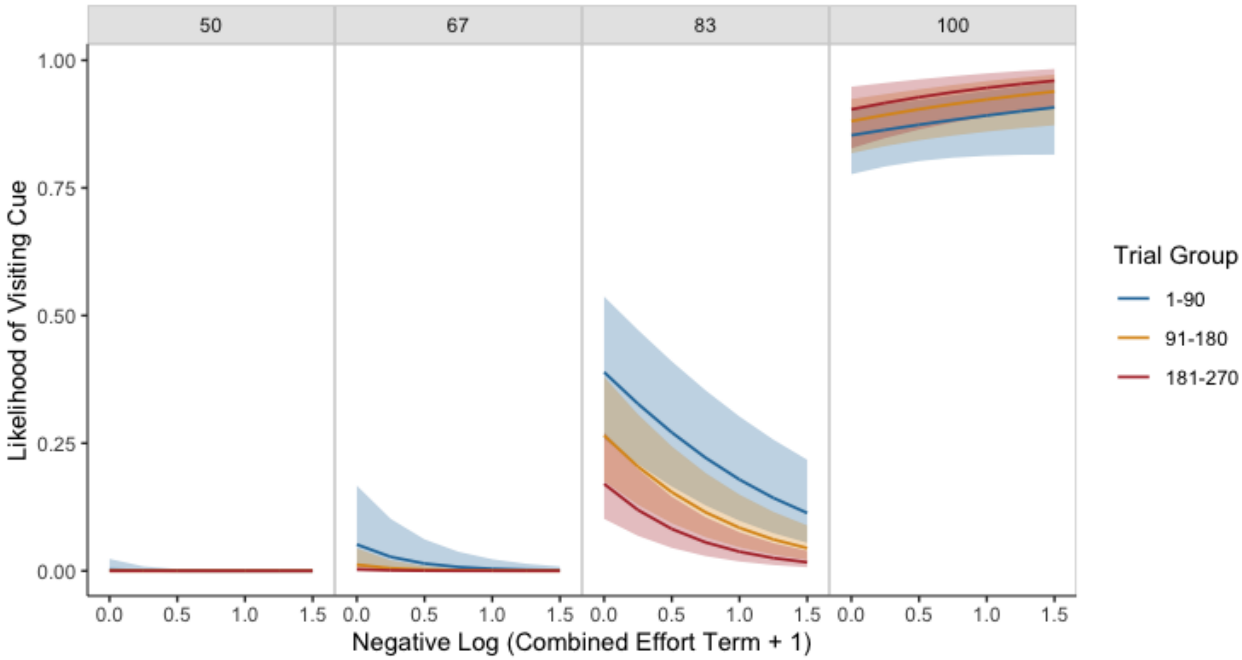
Parameter Estimates for Logistic Regression Predicting Likelihood of Cue Visits for Experiment 1

	B	*SE*	*z*	*p*
Intercept	−5.05	1.31	−3.87	<.001
Effort	−4.26	0.82	−5.21	<.001
Predictive Validity	9.79	2.10	4.67	<.001
Trial	−8.31	2.19	−3.80	<.001
Effort × Predictive Validity	6.65	1.47	4.51	<.001
Effort × Trial	−4.39	0.56	−7.83	<.001
Predictive Validity × Trial	13.22	3.74	3.53	<.001
Effort × Predictive Validity × Trial	7.03	0.85	8.32	<.001

Note. Parameter estimates for the multilevel logistic regression predicting likelihood of cue visits. The parameter of "Effort" is made up of −1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of Trial is made up of trial number / 270. This model included random intercept and slope effects of effort, predictive validity, trial, and all possible two-way interactions. Estimates are given in log odds space.

Figure 41

Conditional Effects of Frequentist Logistic Regression for Experiment 1



Note. In this figure, the Effort variable appears as it did in the analyses, log transformed (log(x+1)), minimum-centered, and multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Confidence Intervals.

Appendix F - Frequentist Models, Categorical PV, Experiment 1

Table 15

Model Comparisons for Logistic Regressions with Predictive Validity as a Categorical Variable Predicting Likelihood of Cue Visits for Experiment 1

Model	Random Effects Syntax	AIC
1. Intercept Only	(1 | Participant)	12448.13
2. Only Main Effects	(1 + Effort + Predictive Validity + Trial | Participant)	10625.89
3. Main Effects and Two-Way Interactions	(1 + Effort + Predictive Validity + Trial + Effort : Predictive Validity + Effort : Trial + Predictive Validity : Trial | Participant)	10468.71
4. Main Effects and All Interactions	(1 + Effort × Predictive Validity × Trial | Participant)	10539.25

Note. Model comparisons for multilevel logistic regressions with Predictive Validity as a categorical variable predicting likelihood of cue visits for Experiment 1. Models only differed in their random effects. The fixed effects Effort, Predictive Validity, Trial, and their interactions, were kept the same for all models. The parameter of "Effort" is made up of −1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of Trial is made up of trial number / 270. Models 3 and 4 were run with the nloptr optimizer to aid with convergence. The R syntax for each random effect structure is included. With the lowest AIC value, Model 3 was chosen as the best fit. For comparison, the best fitting model with Predictive Validity as a continuous predictor had an AIC of 14504.86.

Table 16

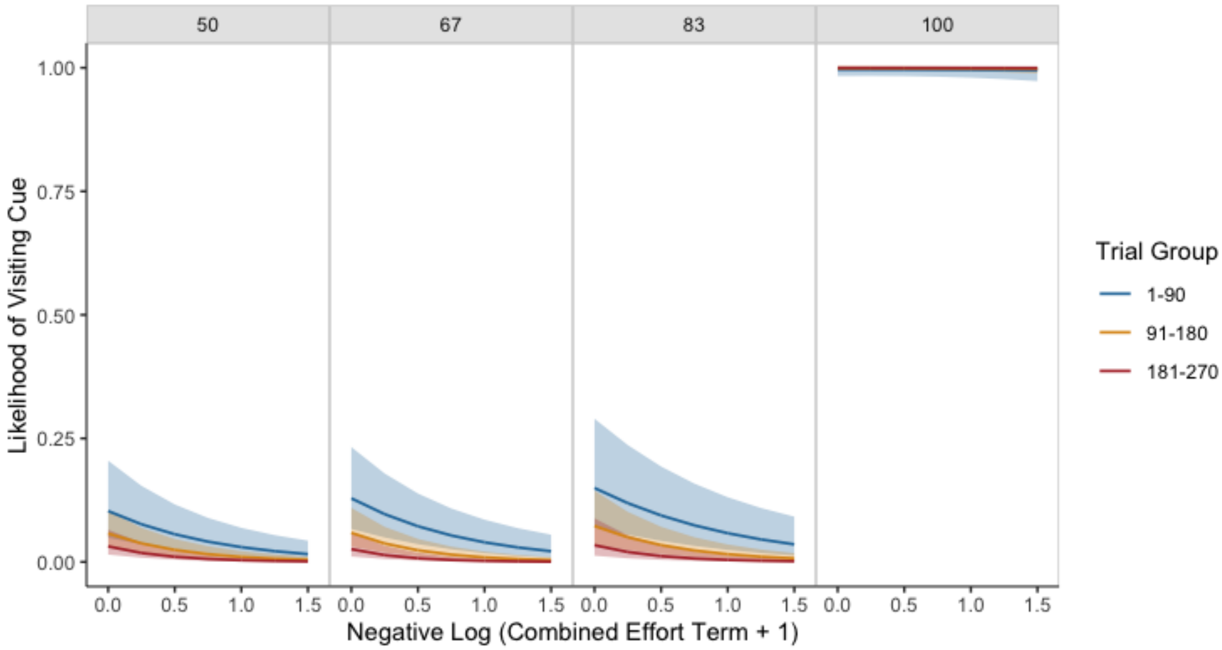
Parameter Estimates for Logistic Regression with Predictive Validity as a Categorical Variable Predicting Likelihood of Cue Visits for Experiment 1

	B	*SE*	*z*	*p*
Intercept	−1.86	0.49	−3.80	<.001
Effort	−1.11	0.33	−3.37	<.001
Predictive Validity [67]	0.36	0.28	1.29	.20
Predictive Validity [83]	0.51	0.29	1.75	.08
Predictive Validity [100]	6.62	0.94	7.04	<.001
Trial	−1.90	0.72	−2.66	.008
Effort: Predictive Validity [67]	0.10	0.30	0.32	.75
Effort: Predictive Validity [83]	0.29	0.29	1.00	.32
Effort: Predictive Validity [100]	1.17	0.60	1.96	.05
Effort:Trial	−1.39	0.61	−2.26	.02
Predictive Validity [67]:Trial	−0.69	0.58	−1.18	.24
Predictive Validity [83]:Trial	−0.53	0.56	−0.95	.34
Predictive Validity [100]:Trial	8.21	2.33	3.52	<.001
Effort:Predictive Validity [67]:Trial	−0.49	0.54	−0.91	.36
Effort: Predictive Validity [83]:Trial	−0.26	0.50	−0.51	.61
Effort: Predictive Validity [100]:Trial	−0.37	0.87	−0.43	.67

Note. Parameter estimates for the multilevel logistic regression predicting likelihood of cue visits. The parameter of "Effort" is made up of −1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The labels of the Predictive Validity levels are written in their original scale (67, 83, 100). The categorical variable of Predictive Validity is dummy coded, with the minimum level 50 as the reference level. The parameter of Trial is made up of trial number / 270. This model included random intercept and slope effects of effort, predictive validity, trial, and all possible two-way interactions. Estimates are given in log odds space.

Figure 42

Conditional Effects of Frequentist Logistic Regression with Categorical PV for Experiment 1

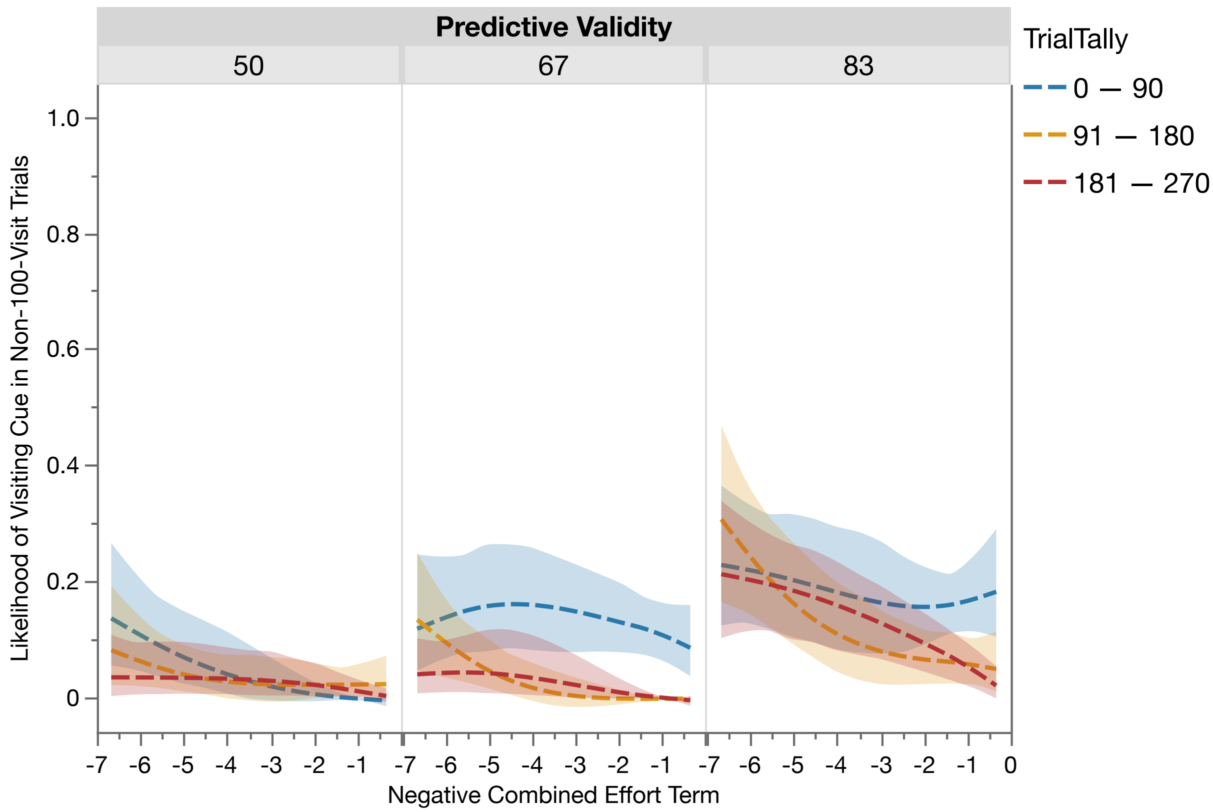


Note. In this figure, the Effort variable appears as it did in the analyses, log transformed (log(x+1)), minimum-centered, and multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Confidence Intervals.

Appendix G - Non-100 Cue Visit Trials, Experiment 1

Figure 43

Raw Data for Cue Visits on Trials Where the 100 Cue was not Visited, Experiment 1

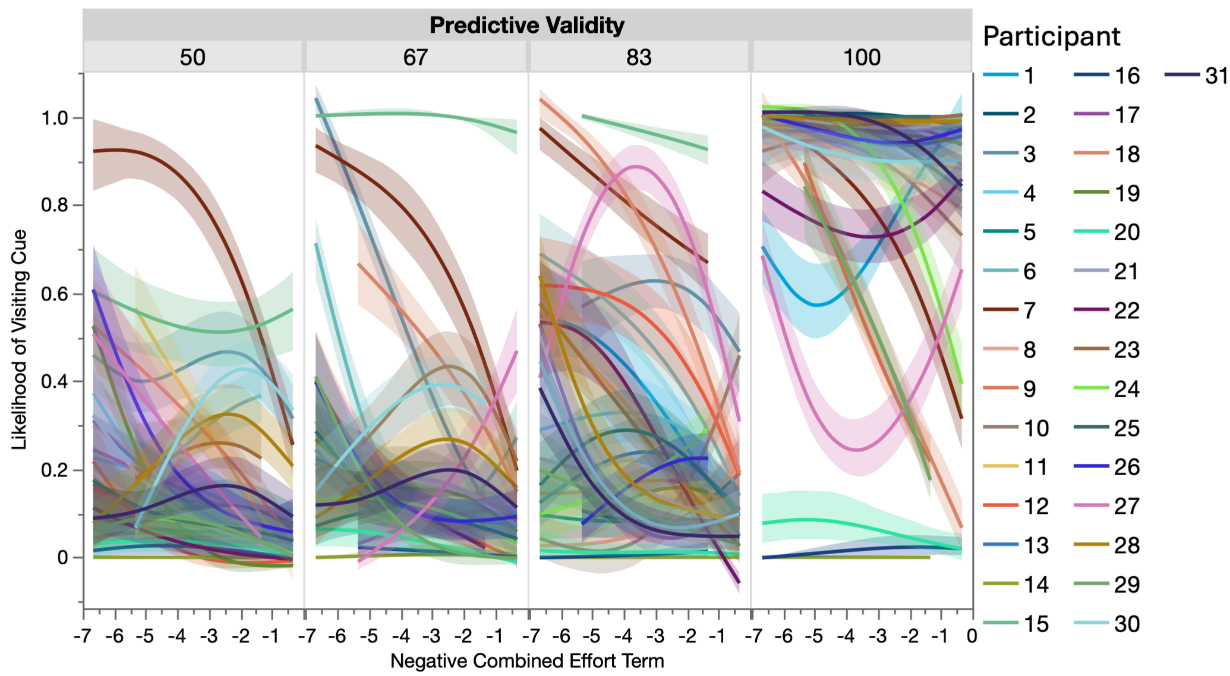


Note. This figure shows the 609 instances where the 100% Predictive Validity cue was not visited, out of 8,100 total trials from all participants (30 participants × 270 trials). As all original responses are 0 (cue not visited on a trial) or 1 (cue visited on a trial), raw data is visualized here with a cubic spline fit (λ = 0, i.e., no smoothing; JMP Student Edition 18.2.1). Error ribbons represent JMP's bootstrapped confidence region for the spline fits.

Appendix H - Additional Raw Data Visualizations for Experiment 2

Figure 44

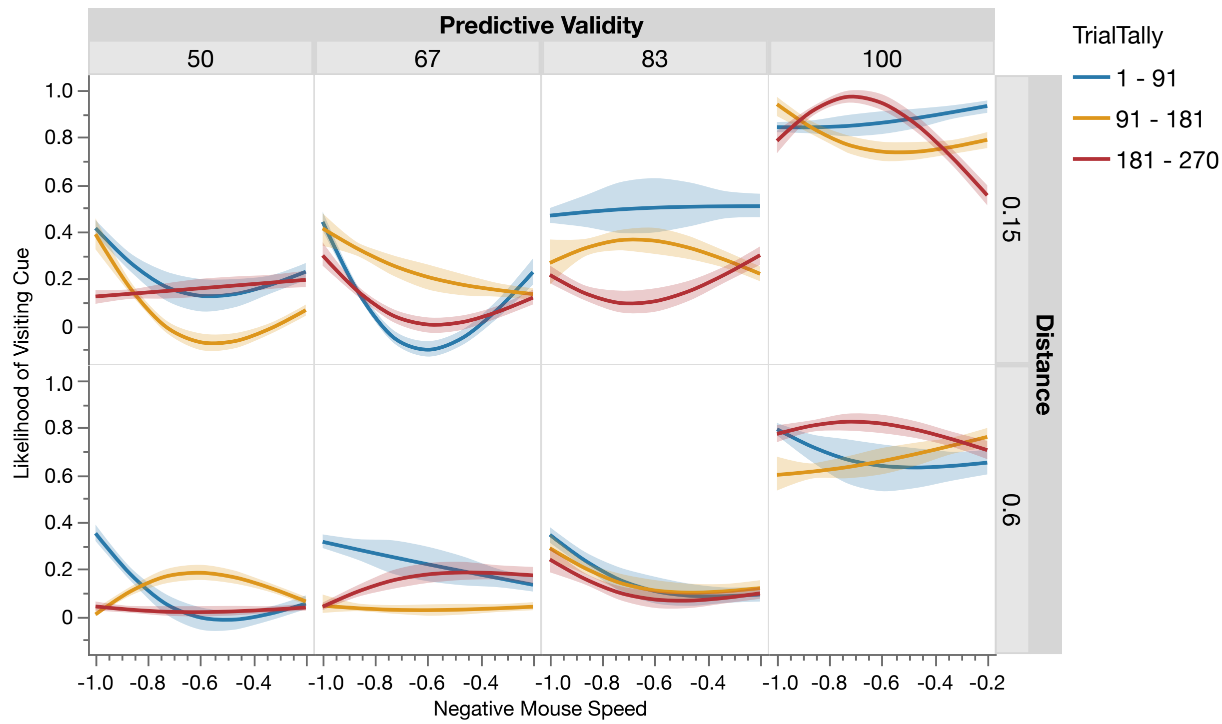
Raw Data for Experiment 2 by Effort, Predictive Validity, and Participant



Note. As all original responses are 0 (cue not visited on a trial) or 1 (cue visited on a trial), raw data is visualized here with a cubic spline fit (λ = 22 to reduce excessive curvature and predictions above 1 and below 0; JMP Student Edition 18.2.1). Any predicted values above 1 and below 0 reflect the spline's interpolation rather than possible likelihoods in the data. Error ribbons represent JMP's bootstrapped confidence region for the spline fits.

Figure 45

Raw Data for Experiment 2 by Mouse Speed, Distance, Predictive Validity, and Trial.

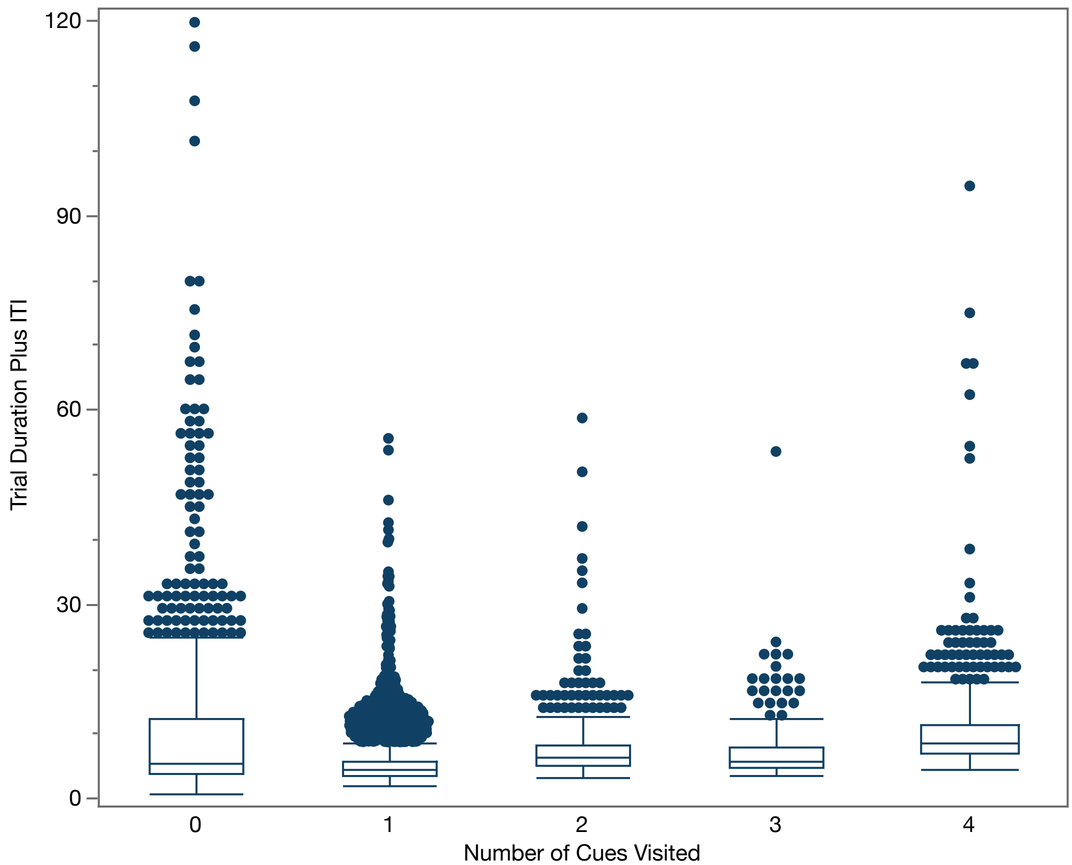


Note. In this figure the combined effort term is split into its component parts of mouse speed coefficient (multiplied by −1 such that lower, more effortful speeds reflect higher values) and distance (in proportion of screen height). This visualization shows no large discrepancies in responses that would caution against the usage of a combined effort term. As all original responses are 0 (cue not visited on a trial) or 1 (cue visited on a trial), raw data is visualized here with a cubic spline fit (λ = 0, i.e., no smoothing; JMP Student Edition 18.2.1). Error ribbons represent JMP's bootstrapped confidence region for the spline fits.

Appendix I - Trial Duration Visualizations for Experiment 2

Figure 46

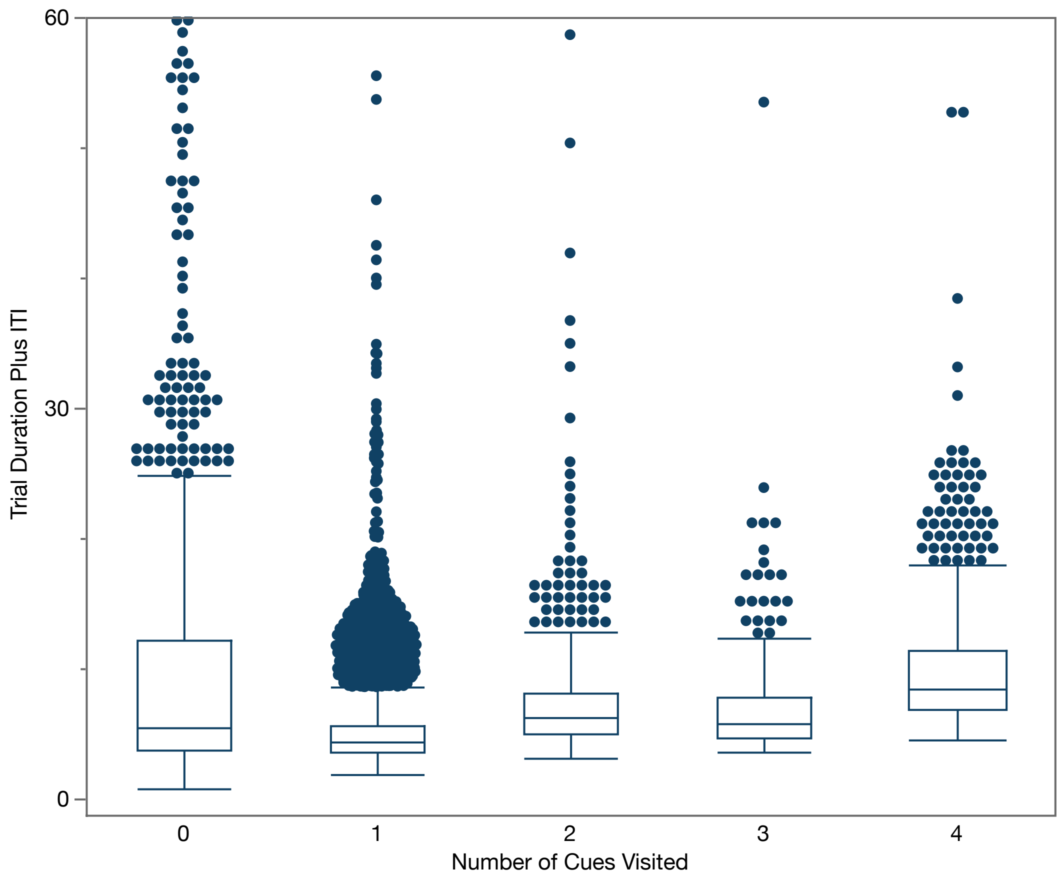
Trial Duration (Including ITI) by Number of Cues Visited for Experiment 2



Note. This figure shows all 8370 trials experienced by participants (31 participants with 270 trials each), minus 3 trials with durations longer than 120 s.

Figure 47

Trial Duration (Including ITI) by Number of Cues Visited for Experiment 2, Zoomed In



Note. This figure shows all 8370 trials experienced by participants (31 participants with 270 trials each), minus 24 trials with durations longer than 60 s.

Table 17

Descriptive Statistics for Trial Duration (Including ITI)

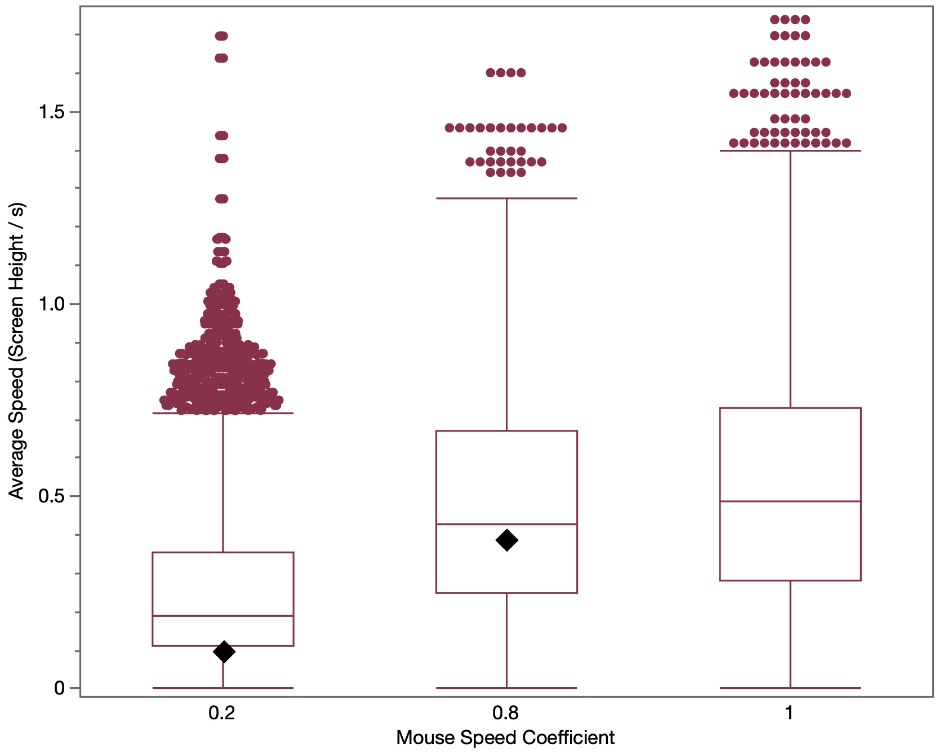
by Number of Cues Visited for Experiment 2

Number of Cues Visited	Mdn	IQR
0	5.48	[3.81, 12.26]
1	4.39	[3.66, 5.65]
2	6.35	[4.98, 8.13]
3	5.85	[4.77, 7.79]
4	8.54	[6.92, 11.38]

Appendix J - Average Speed Per Mouse Setting, Experiment 2

Figure 48

Trial Average Speed by Mouse Speed Coefficient Setting for Experiment 2

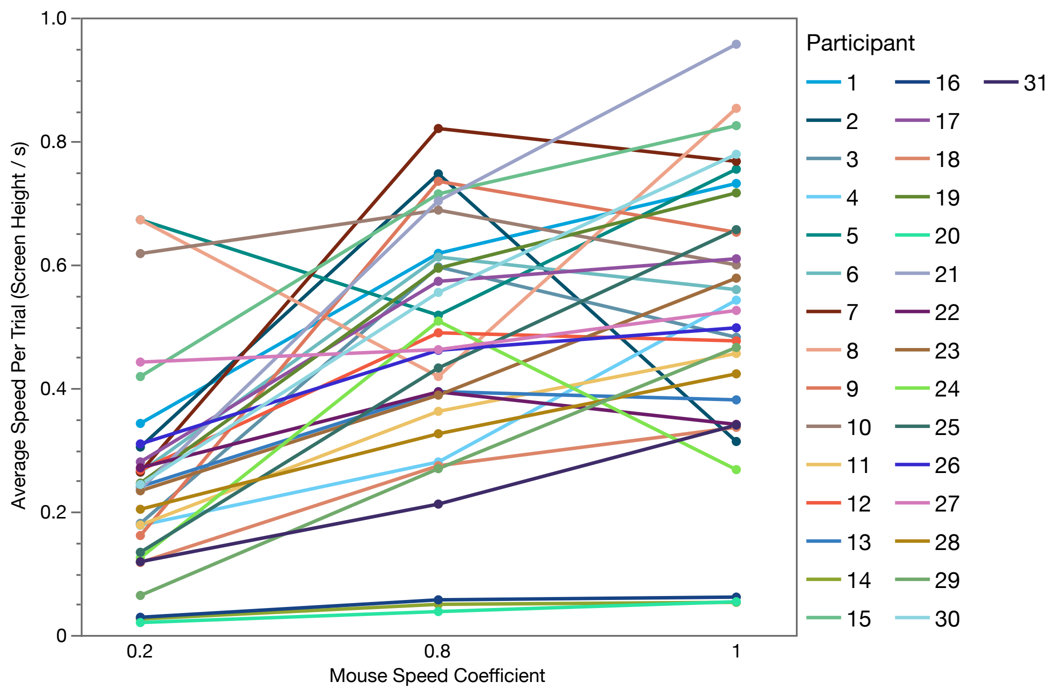


Note. This figure shows the average speed in screen height units per second for the 8370 trials experienced by participants (31 participants with 270 trials each). When the mouse speed was unchanged from the participants' normal speed, participants averaged a median of 0.49 screen height / s (*IQR* = [0.28, 0.73]). Diamonds on the plots for the 0.2 and 0.8 speed coefficients represent the anticipated outcome speeds relative to this 100% speed outcome (80% × 0.49 = 0.39; 20% × 0.49 = 0.10). When the mouse speed was set to 80% of its original speed, participants moved at a median speed of 0.43 screen height / s (*IQR* = [0.25 , 0.67]; 88% of the median for 100% speed). This faster outcome speed is possible by moving the mouse or trackpad physically more in a shorter amount of time than was done at the 100% speed setting.

Participants appeared to make even larger physical movements to compensate for the slowest mouse speed of 20%, with participants moving a median of 0.19 screen height / s (*IQR* = [0.11 , 0.35]; 39% of the median for 100% speed).

Figure 49

Trial Average Speed by Mouse Speed Coefficient Setting for Experiment 2, by Participant



Note. This figure shows the average speed in screen height units per second for each participant (90 trials per speed coefficient per participant).

Appendix K - Frequentist Models, Continuous PV, Experiment 2

Table 18

Model Comparisons for Logistic Regressions Predicting Likelihood of Cue Visits for Experiment 2

Model	Random Effects Syntax	AIC
1. Intercept Only	(1 | Participant)	26353.54
2. Only Main Effects	(1 + Effort + Predictive Validity + Trial + Block| Participant)	22783.15

Note. Model comparisons for multilevel logistic regressions predicting likelihood of cue visits for Experiment 2. Models only differed in their random effects. The fixed effects Effort, Predictive Validity, Trial, and their interactions, were kept the same for all models. The parameter of "Effort" is made up of −1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of Trial is made up of trial number / 270. The R syntax for each random effect structure is included. All models with interactions in the random effects structure did not converge with bobyqa, nloptr, or Nelder-Mead optimizers. With the lowest AIC value of the converging models, Model 2 is the best fit.

Table 19

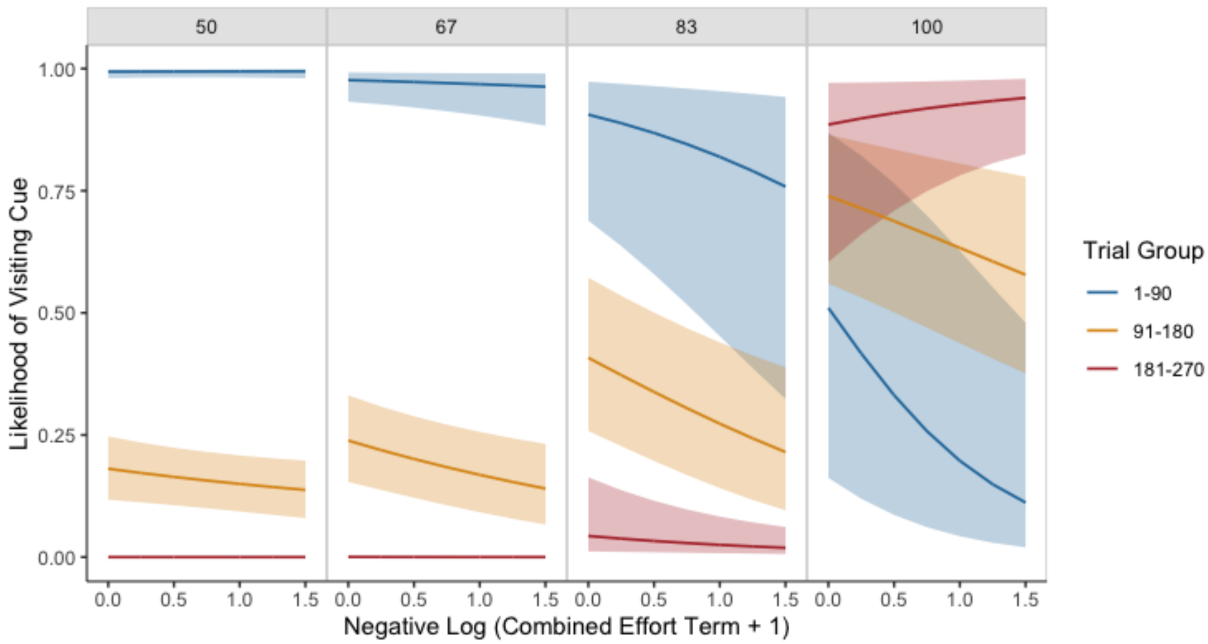
Parameter Estimates for Logistic Regression Predicting Likelihood of Cue Visits for Experiment 2

	B	*SE*	*z*	*p*
Intercept	-1.29	0.73	-1.75	.07
Effort	-3.49	0.36	-9.76	< .0001
Predictive Validity	5.54	0.82	6.78	< .0001
Trial	-47.69	3.15	-15.14	< .0001
Block	5.04	0.47	10.76	< .0001
Effort × Predictive Validity	3.99	0.63	6.31	< .0001
Effort × Trial	-5.96	1.96	-3.04	.002
Predictive Validity × Trial	71.24	3.83	18.62	< .0001
Effort × Block	1.67	0.30	5.62	< .0001
Predictive Validity × Block	-8.61	0.58	-14.80	< .0001
Trial × Block	1.77	0.18	9.78	< .0001
Effort × Predictive Validity × Trial	14.84	3.73	3.97	< .0001
Effort × Predictive Validity × Block	-3.00	0.57	-5.24	< .0001
Effort × Trial × Block	-0.37	0.19	-1.89	.06
Predictive Validity × Trial × Block	-1.72	0.37	-4.63	< .0001
Effort × Predictive Validity × Trial × Block	0.15	0.39	0.39	.70

Note. Parameter estimates for the multilevel logistic regression predicting likelihood of cue visits for Experiment 2. The parameter of "Effort" is made up of −1(log(Effort + 1) centered by its minimum. The parameter of "Predictive Validity" is made up of log(predictive validity + 1) centered by its minimum. The parameter of Trial is made up of trial number / 270. This model included random intercept and slope main effects of effort, predictive validity, trial, and block. Estimates are given in log odds space.

Figure 50

Conditional Effects of Logistic Regression for Experiment 2, with Predictive Validity as Continuous Variable

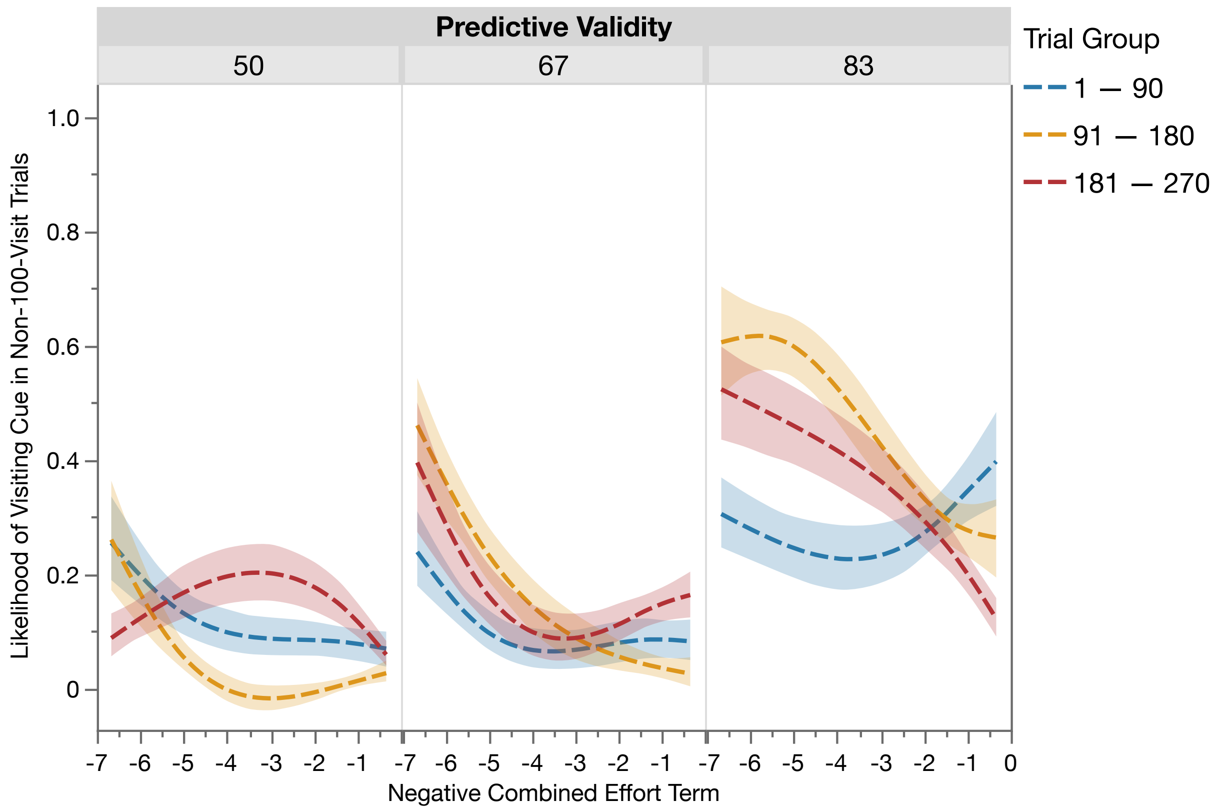


Note. In this figure, the Effort variable appears as it did in the analyses, log transformed (log(x+1)), minimum-centered, and multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Error ribbons represent 95% Confidence Intervals.

Appendix L - Non-100 Cue Visit Trials, Experiment 2

Figure 51

Raw Data for Cue Visits on Trials Where the 100 Cue was not Visited, Experiment 2

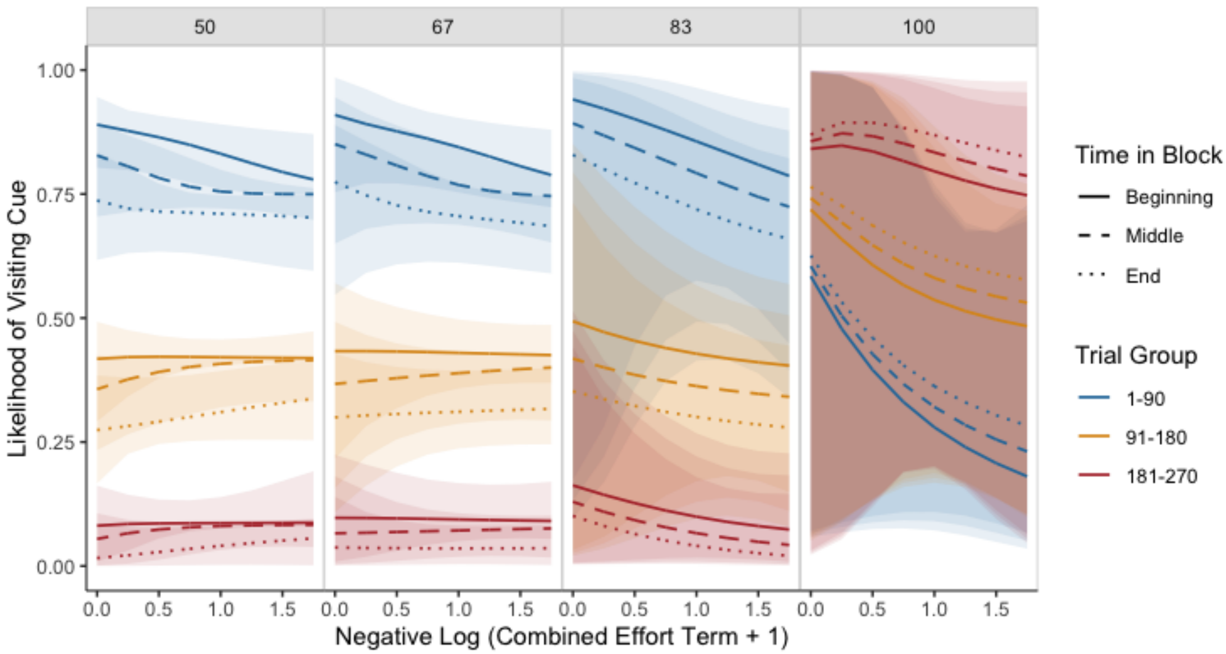


Note. This figure shows the 1,929 instances where the 100% Predictive Validity cue was not visited, out of 8,370 total trials from all participants (31 participants × 270 trials). As all original responses are 0 (cue not visited on a trial) or 1 (cue visited on a trial), raw data is visualized here with a cubic spline fit (λ = 0, i.e., no smoothing; JMP Student Edition 18.2.1). Error ribbons represent JMP's bootstrapped confidence region for the spline fits.

Appendix M - Conditional Effects for Experiment 2, by Block

Figure 52

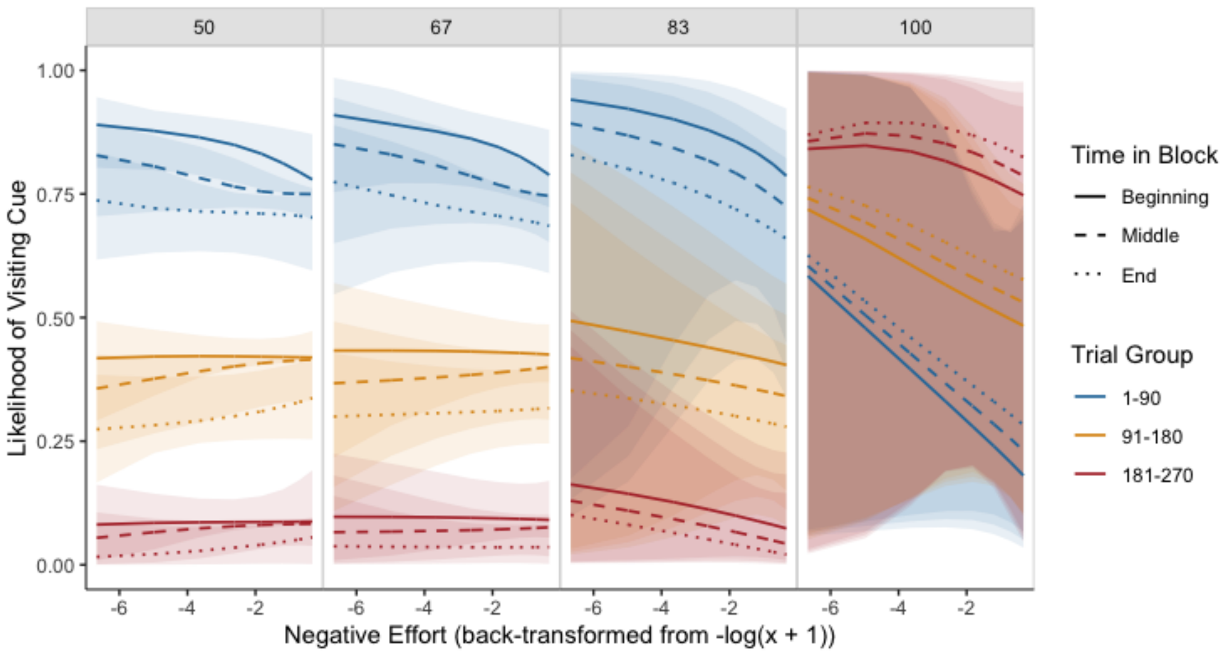
Conditional Effects of Logistic Regression for Experiment 2, Continuous PV, Separated by Block



Note. In this figure, the Effort variable appears as it did in the analyses, log transformed (log(x+1)), minimum-centered, and multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Time in Block represents the Beginning 15 trials, Middle 15 trials, and End 15 trials of all blocks. Error ribbons represent 95% Credible Intervals.

Figure 53

Conditional Effects of Logistic Regression for Experiment 2, Continuous PV, Separated by Block, Back Transformed



Note. In this figure, the Effort variable is back-transformed to its original scale but still multiplied by −1 to ease in interpretation (causing the effort variable to increase instead of decrease with higher difficulty). Predictive Validity and Trial labels reflect their original, non-transformed values. Time in Block represents the Beginning 15 trials, Middle 15 trials, and End 15 trials of all blocks. Error ribbons represent 95% Credible Intervals.