AN ACCESSIBILITY FRAMEWORK FOR CUE-BASED INFERENCES

A Dissertation
Presented to
The Academic Faculty

By

Ashley Lawrence-Huizenga

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Psychology

Georgia Institute of Technology

December 2018

An Accessibility Framework for Cue-Based Inferences

Approved by:

Dr. Rick Thomas, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Frank Durso
School of Psychology
Georgia Institute of Technology

Dr. Mark Wheeler
School of Psychology
Georgia Institute of Technology

Dr. Daniel Spieler
School of Psychology
Georgia Institute of Technology

Dr. Paul Elmore
Geospatial Sciences and Technology Branch
Naval Research Laboratory

Date Approved: November 7, 2018

ACKNOWLEDGMENTS

I would like to acknowledge and thank the following people who have supported me during my graduate career. First, I would like to thank my graduate advisor, Dr. Rick Thomas, for his guidance and encouragement on this dissertation. Second, I would like to thank my dissertation committee members, Dr. Frank Durso, Dr. Mark Wheeler, Dr. Dan Spieler, and Dr. Paul Elmore, for their helpful insights. Finally, I would also like to thank my friends and family for their support and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Many studies throughout the area of decision-making have shown that people are able to adapt to different decision environments. A number of frameworks have been proposed that seek to explain adaptive decision making in the context of cue-based inferences, a type of decision where a person decides which option is highest on a variable of interest based on the attributes of those options. However, current frameworks fail to account for the role of memory in cue-based inferences. The goal of this dissertation was to test whether a framework based on the accessibility of cues in memory can provide a better account of adaptive decision-making in cue-based inferences compared to either the adaptive toolbox or current single-strategy models. Three experiments were conducted to test the accessibility framework by manipulating decision environments as well as directly manipulating memory for cues. The results of the experiments extend previous research showing that memory affects cue-based inferences, challenging frameworks that are based on validity only. They also extend research on adaptive decision-making by showing that people are sensitive to the decision environment but that this does not always result in changes to both decision outcomes and decision processes. Overall, the accessibility framework provides a promising foundation for explaining how people make cue-based inferences, but further research is necessary to better understand how people search for cues, particularly how they decide to stop searching.

CHAPTER 1: INTRODUCTION

People are often faced with decisions in which they are trying to choose which option is highest (or lowest) on some variable of interest based on the attributes of those options. For example, a restaurant-goer might want to choose which menu item has fewer calories based on different characteristics of the menu items, such as whether they are made with red meat, whether they have dairy, whether they are high on carbs, etc. How a person makes their decision might change depending on the decision environment. For example, which attributes they use and how they combine those attributes to arrive at a decision may depend on how much time they have to make the decision and how easy the attribute information is to retrieve from the environment or from memory. There is a large body of literature that seeks to understand how people make these inferences and how people adapt their decision-making to the environment (for review see Gigerenzer, Hertwig, & Pachur, 2011). Within the literature, these inferences are known as cue-based inferences because people are using information about the options to infer the value of a criterion of interest, which is calories in the above example. Information about the options that can be used to make inferences are known as cues, such as the presence of dairy for the menu items being compared.

There is ample evidence that people making cue-based inferences adapt to the decision environment by either using fewer cues or combining the cue information differently (Bröder, 2003; Newell & Shanks, 2003; Payne, Bettman, & Johnson, 1988; Rieskamp & Hoffrage, 2008; Rieskamp & Otto, 2006). Much of the work within cue-based inferences is based on the work of Payne et al. (1988), who proposed the idea that people select adaptively among different strategies based on the decision environment. These strategies are characterized by rules for searching cues, when to stop searching, and how to combine cues. The most popular account of adaptive decision making within the cue-based inference literature is the adaptive toolbox (Gigerenzer, Todd, & ABC Research Group, 1999). The adaptive toolbox, like the theory of adaptive decision making (Payne et al., 1988), is a multi-strategy framework that assumes people select adaptively among a repository of qualitatively different strategies depending on the decision environment. Much of this research compares under which circumstances people are likely to use compensatory strategies, high values on less useful cues can compensate for a low value on a highly useful cue, versus noncompensatory strategies, only the most useful cue is used. Recently, however, this framework has been criticized, both because the strategy selection process is poorly specified (Glöckner & Betsch, 2008a; Lee & Cummins, 2004; Newell, 2005; Söllner, Bröder, Glöckner, & Betsch, 2014) and because the increasing number of strategies in the toolbox makes it difficult to falsify (Dougherty, Thomas, & Franco-Watkins, 2008; Marewski & Link, 2014; Scheibehenne, Rieskamp, & Wagenmakers, 2013).

Single-strategy frameworks have been proposed as potential solutions to the problems with multi-strategy accounts. The single-strategy frameworks, such as evidence accumulation (Lee & Cummins, 2004) and parallel constraint satisfaction (Glöckner & Betsch, 2008a), assume that people use one strategy that adjusts to the decision environment. For example, the threshold of the amount of evidence required to make a decision adjusts to the decision environment in the evidence accumulation model (Lee & Cummins, 2004). Single-strategy frameworks can account for much the same data as the multi-strategy frameworks because they are able to mimic apparent strategy changes by adjusting thresholds or information weights (Glöckner, Betsch, & Schindler, 2010; Glöckner & Hodges, 2011; Lee & Cummins, 2004; Newell & Lee, 2011).

Although there is evidence for both the multi-strategy and single-strategy frameworks, none of the current frameworks provide a fully compelling account of adaptive decision-making in cue-based inferences. Most frameworks rely on the concept of pre-computed cue hierarchies based on cue validity, a measure of cue accuracy. Specifically, cue validity is the number of times a cue correctly identifies the correct option compared to the total number of times it differs between the options. The adaptive toolbox assumes cue use is driven by cue validity and evidence accumulation models assume search follows validity order. In these frameworks, cue use is based on the cues being organized hierarchically such that cues are used in order of validity with the hierarchy being determined prior to decision-making (precomputed). The general concept of a precomputed hierarchy is not psychologically plausible because of the amount of knowledge required to calculate cue validity and the number of calculations that need to be made (Dougherty, Franco-Watkins, & Thomas, 2008). Those frameworks that are not directly based on validity, such as parallel constraint satisfaction, still do not provide compelling accounts of decision-making because they do not fully specify how people search and select which cues are used in the decision process. Parallel constraint satisfaction tries to separate itself from the formal definition of cue validity by invoking subjective cue validity. Subjective cue validity is determined by the decision maker and does not always reflect actual cue validity. However, there is no explicit explanation of the processes involved in determining subjective cue validity (Glöckner & Betsch, 2008a).

More importantly, all of these frameworks ignore the central role of memory in cue-based inferences. There is evidence that accessibility affects cue use beyond what can be accounted for by validity alone (Lawrence, Thomas, & Dougherty, 2018a, 2018b; Platzer & Bröder, 2012; Platzer, Bröder, & Heck, 2014;). Within in this dissertation, accessibility refers to how easily information can be retrieved from memory. For example, people tend to use more accessible cues even when they have low validities (Lawrence et al., 2018a, 2018b; Platzer et al., 2014). There is also evidence that cue accessibility affects later decision strategies such that people are more likely to use compensatory strategies when less valid cues are accessible and more likely to use noncompensatory strategies when more valid cues are accessible (Lawrence et al., 2018a; Platzer et al., 2014). A good explanation of cue-based inferences should be able to account for these effects.

The goal of this dissertation is to propose and test a framework for cue-based inferences that is based on memory accessibility. Because this framework focuses on the role of memory in cue-based inferences, it can account for accessibility effects that the other frameworks cannot. Moreover, the framework can account for apparent changes in strategy via accessibility, thus, avoiding the strategy selection problem. Unlike the other single-strategy frameworks, the use of cues is described in a psychologically plausible manner that is not based on a validity hierarchy but is instead based on well-established memory phenomenon. Overall, an accessibility-based framework should lead to a more parsimonious account of the data and a testable model of decision making.

A series of experiments were conducted to test the accessibility framework. Previous experiments have shown that cue accessibility affects how cues are used such that more accessible cues are preferred over less accessible ones (Lawrence et al., 2018a, 2018b). However, these experiments focused on cue preference and failed to demonstrate that an accessibility-based framework truly provides the best account of cue-based inferences in terms of both cue search and decision process. In order to fully test the accessibility framework in this dissertation, experiments were conducted that manipulated decision tasks in ways that commonly lead to different apparent strategies to show that an accessibility-based framework can account for these changes. For example, cue dispersion, the amount of variability in the validities of the cues, has been shown to affect apparent decision strategies (Bröder, 2003; Newell & Shanks, 2003) and the accessibility account should be able to explain the effects of cue dispersion. Moreover, the experiments also manipulated accessibility via well-established memory effects to show that the framework can also account for accessibility effects beyond what current frameworks can account for. In general, these experiments tested how direct manipulations of accessibility interact with decision environment.

Experiment 1 extended previous research which found evidence for serial position effects on cue preference (Lawrence et al., 2018b). Specifically, this experiment tested the influence of accessibility on cue preference and decision by crossing serial position manipulations with manipulations of average cue validity and dispersion of cue validity. Experiment 2 sought to generalize the effects of accessibility to a manipulation of accessibility via a retention interval while still maintaining the manipulation of cue dispersion. Moreover, this study provided more detailed process data than the previous experiment by allowing participants to decide when to stop searching on their own. Experiment 3 then tested whether an accessibility framework could also account for decisions made from information stored in memory. One of the major arguments from supporters of the adaptive toolbox account is that people are more likely to use noncompensatory strategies when the decisions are made from memory rather than from information on the screen (Bröder & Gaissmaier, 2007; Bröder & Schiffer, 2003). Thus, it is important to show that an accessibility account can also function in this context.

The data from these experiments were compared to behavior predicted from an accessibility-based framework. The framework evaluated was an extension of the HyGene architecture (Lawrence et al., 2018a; Thomas, Dougherty, Sprenger, & Harbison, 2008). Specifically, the model assumes that cue preference is driven by the accessibility of the cues in memory, and it assumes that the cues used to make a decision are limited to only those available in working memory. The model mimics apparent strategy changes through changes in which cues are accessible to the decision-maker. Those cues that are accessible are assumed to be integrated in a compensatory manner to arrive at a decision. The goal of the analyses was to test the predictions made by the accessibility model in terms of cue search and final decision by comparing those predictions to participant behavior. The predictions of the accessibility model were also compared to the predictions of a general evidence accumulation model (Lee & Cummins, 2004), parallel constraint satisfaction (Glöckner & Betsch, 2008a), and predictions based on the adaptive toolbox.

The experiments further the understanding of how people make cue-based inferences by testing the effect of the interaction between direct manipulations of accessibility and manipulations of the decision environment. The experiments partially establish that accessibility can account for both cue use and decisions beyond what can be accounted for by currently available frameworks. Testing the accessibility account is important because it is more plausible and parsimonious than other accounts of cue-based inferences. It is more plausible in that it assumes that cue use is based on well-established memory phenomenon rather than a pre-computed validity hierarchy. Because it does not require a vast number of free parameters to account for behavior, it is also more parsimonious than the other frameworks. By extending the understanding of the importance of memory in how people make cue-based inferences, this work provides a major contribution to the literature.

The rest of the paper will be structured as follows. Literature concerning the current accounts of cue-based inferences will be reviewed, focusing on evidence for and against these frameworks for adaptive decision-making. Evidence for the role of memory in cue-based inferences will then be discussed to highlight some of the shortcomings of these current accounts. A detailed description of the accessibility-based framework will be provided as an alternative account for the adaptive decision-making in cue-based inferences. Then three experiments to test this account will be discussed. The paper will conclude by highlighting the major findings from these experiments and discussing further research.

CHAPTER 2: LITERATURE REVIEW

2.1 Multi-Strategy Framework

The most researched account of adaptive decision-making in cue-based inferences is the adaptive toolbox (Gigerenzer et al., 1999). This framework assumes that people adapt to changes in the environment by selecting among qualitatively different strategies. Each strategy in the toolbox is described by a search rule, a stopping rule, and a decision rule (Gigerenzer & Goldstein, 1996). There are numerous strategies that are included within the adaptive toolbox (for review see Gigerenzer et al., 2011). This dissertation will not review all the strategies within in the adaptive toolbox but will focus on those strategies that search through cue information for the options and decide based on that information.

Strategies can be either compensatory or noncompensatory. In compensatory strategies, positive values on some cues can compensate for negative values on other cues. Alternatively, in noncompensatory strategies, decisions are based on a single cue that cannot be compensated for by other cues. Much of the work within the adaptive toolbox builds off of the work of Payne, Bettman, and Johnson (for review see Payne, Bettman, & Johnson, 1993) who studied how decision makers adapt to different decision environments. Although Payne et al.'s (1993) research takes a different theoretical approach than the adaptive toolbox research, their focus on how people adapt to decisions environments provided the base for the adaptive toolbox research. They found that how people process information while making a decision is affected by both the probability dispersion of the options and time pressure (Payne et al., 1988; Payne, Bettman, & Luce, 1996). Specifically, they found that high dispersion and high time pressure lead to more attribute-based processing compared to alternative-based processing. Importantly, attribute-based processing is often considered to be a characteristic of noncompensatory strategies and alternative-based is often considered to be a characteristic of compensatory strategies. In general, evidence that decision-makers are sensitive to the decision environment and adaptively adjust between compensatory strategies and noncompensatory strategies supports the adaptive toolbox framework.

The most popular and well researched noncompensatory strategy for cue-based inferences within the toolbox is the Take-the-Best (TTB) heuristic (Gigerenzer & Goldstein, 1996). The search rule assumes that decision makers search available cues by their validity, where validity is defined as the proportion of times a cue correctly predicts which option is higher on the criterion given each option has a different value for that cue (the cue discriminates). Options are compared for the most valid cue first. If this cue does not discriminate between options, then the options are compared for the next most valid cue and so on. Once a discriminating cue is found, decision makers stop searching cues and choose the option that is higher on that cue. If no discriminating cues are found, then the decision maker guesses. This strategy has often been compared to a weighted additive strategy (WADD), which is a compensatory strategy that integrates all cue information (values and validities) by weighting cues by their validity and summing.

Work specifically testing the adaptive toolbox in cue-based inferences have also found that people behave adaptively. Increasing the cost of information results in the use of noncompensatory strategies rather than compensatory ones. For example, people are more likely to use noncompensatory strategies when the cues are assigned a high monetary cost (Bröder & Gaissmaier, 2007; Bröder & Schiffer, 2003). This also occurs when information has a retrieval cost such that people are more likely to use noncompensatory strategies when information must be retrieved from memory (Bröder, 2000; Newell & Shanks, 2003). Increasing time pressure also results in the use of noncompensatory strategies rather than compensatory ones (Rieskamp & Hoffrage, 2008), but this may depend on whether the information is presented sequentially or simultaneously (Glöckner & Betsch, 2008b). The dispersion of cue validities also affects strategy use such that when cues are highly dispersed, meaning one cue is much more valid than the others, people are more likely to use a noncompensatory strategy (Bröder, 2003; Newell & Shanks, 2003). In general, participants adapt strategies based on the ecological structure of the environment such that they seem to use the strategy that performs the best in that environment (Bröder, 2003; Rieskamp & Hoffrage, 2008; Rieskamp & Otto, 2006; Lawrence et al., 2018a).

Despite evidence that people seem to adaptively select among strategies, there are a number of criticisms of the adaptive toolbox. Several researchers have expressed concern in regards to the poor specification of the strategy selection process (Glöckner & Betsch, 2008a; Lee & Cummins, 2004; Newell, 2005; Söllner et al., 2014). Within the adaptive toolbox, it is unclear exactly what processes are involved in selecting among the qualitatively different strategies, which some argue leads to a recursive process in which the decision maker is stuck in a loop of deciding how to decide (Glöckner & Betsch, 2008a). Attempts have been made to address this problem, for example, Rieskamp and Otto (2006) proposed a reinforcement learning model in which people learn which strategies perform well based on the past performance of those strategies. Although there is some evidence for this model over an exemplar-based model, the model tested assumed only two strategies, WADD and TTB. It is unclear that it would be able to scale up to the number of strategies assumed to be in the adaptive toolbox. Moreover, the model assumes that the toolbox strategies are already represented in the mind without explaining how all of these different strategies were learned in the first place. Marewski and Schooler (2011) also addressed the strategy selection problem by proposing the cognitive niche framework. In this framework, the environment and memory are assumed to limit which strategies can be applied in certain situations. Thus, there can be situations in which only one strategy can be used, eliminating the need for strategy selection. However, it is unclear what processes are involved when more than one strategy can be applied to an inference decision.

Moreover, these attempts to solve the strategy selection problem do not solve other criticisms of the adaptive toolbox. Newell (2005) argued that the deterministic nature of many of the strategies in the toolbox (e.g. TTB), makes it difficult for the framework to account for individual differences. This is particularly problematic because the framework assumes that the environment determines the strategy, but empirical evidence suggests there are large individual differences (Newell, 2005). Further, the multiple strategy structure of the adaptive toolbox means new strategies can always be proposed to account for decision behavior, making the framework difficult to falsify (Dougherty, Thomas, & Franco-Watkins, 2008; Marewski & Link, 2014; Scheibehenne et al., 2013). The number of strategies in the toolbox, even if the individual strategies are simple, results in a very complex and flexible model (Scheibehenne et al., 2013).

Alternatives to the adaptive toolbox have been proposed that attempt to address the problems with strategy selection, falsification, and individual differences.

2.2 Single-Strategy Frameworks

Single-strategy frameworks have been proposed as alternatives to the adaptive toolbox. Instead of assuming that people select among qualitatively different strategies, these frameworks assume a single strategy that adjusts to the decision environment via changes in decision thresholds (Lee & Cummins, 2004) or changes in cue weighting (Glöckner & Betsch, 2008a). The most studied single-strategy frameworks within the cue-based inference literature are evidence accumulation models (Lee & Cummins, 2004) and parallel constraint satisfaction (Glöckner & Betsch, 2008a). Other models, such as decision field theory (Busemeyer & Townsend, 1993), could also be considered single-strategy models but have not been studied in the context of cue-based inferences. The single-strategy frameworks discussed in this dissertation can account for much of the same data as the adaptive toolbox, such as evidence that people use less information under time pressure and information costs (Bröder & Gaissmaier, 2007; Bröder & Schiffer, 2003; Rieskamp & Hoffrage, 2007). However, they also address some of the criticisms of the adaptive toolbox by eliminating the need for strategy selection and allowing for individual differences.

The parallel constraint satisfaction model (PCS) was proposed as a single-strategy alternative to the adaptive toolbox (Glöckner & Betsch, 2008a). Unlike the adaptive toolbox, PCS assumes compensatory and simultaneous processing of all information (cue values and validity) with information being integrated through a consistency maximizing process. Specifically, the model follows three or four steps (Glöckner & Betsch, 2008a; Glöckner & Hodges, 2011). The first step is the activation of information related to the decision problem (cues, goals, options, etc.) both from memory and from the environment. This information is then used to form a mental representation (network) of the problem. The second step is an automatic consistency maximization process in which activations of options and cues within the network are adjusted until one option dominates the other. Importantly, during this process the validities of cues are also adjusted from their starting values to achieve consistency, meaning posterior cue validities may differ from the a priori cue validities of the network. In the third step, once a consistency threshold is met, the decision maker chooses the option that dominates the other within the network. The fourth step only occurs if consistency is below the threshold. In this step, the structure of the network is deliberately changed to feed back into the consistency maximization process. During this step, the decision maker must select a strategy for searching, producing, or changing information within the network to reach the threshold within the second step.

The evidence accumulation model (EAM) was also proposed to unify both noncompensatory and compensatory strategies within a single-strategy framework (Lee & Cummins, 2004; Newell, 2005). Unlike PCS, this model assumes information is considered sequentially rather than simultaneously. This model follows a random walk process in which evidence from each cue updates the state of the random walk sequentially. Specifically, cues are considered in validity order and evidence is updated via a log-odds transformation of the validity. Information accumulates in this way until a decision threshold is reached and the alternative with the most evidence is selected. If the threshold is not reached after all cues are considered, then the option with the most evidence is selected. Noncompensatory strategies such as TTB and compensatory strategies such as WADD correspond to different evidence accumulation thresholds. Noncompensatory strategies have lower thresholds leading to a decision being made after the most valid discriminating cue. Compensatory strategies have higher evidence thresholds requiring all information to be considered before making a decision.

Both PCS and EAM have empirical support in terms of fitting participant decision behaviors. In decisions where there was time pressure and cue information had to be retrieved from memory, Glöckner and Hodges (2011) classified 40% of participants as using PCS. When inferences were made from given information, around 75% of participants were best fit by PCS (Glöckner & Bröder, 2011). Both of these studies suggest that PCS can account for a decent proportion of participants decision behaviors. In studies comparing EAM with toolbox strategies, EAM best fit about 85% of the participants' decisions in environments in which participants learned cue validities through experience (Lee & Cummins, 2004; Newell & Lee, 2011). Similar results were also found both when cues were costly and when they were not (Newell & Lee, 2011). Both of the single-strategy frameworks and the adaptive toolbox typically make the same predictions in terms of decision behavior, making it difficult to provide conclusive support for them over the adaptive toolbox when looking only at decisions.

However, some attempts have been made to disentangle the single-strategy frameworks from the adaptive toolbox. Because PCS operates through consistency maximization, it assumes that decisions with initially inconsistent information both take longer and result in lower confidence than decisions with initially consistent information.

EAM is also sensitive to consistency such that information inconsistent with the currently favored option moves the decision maker further from the threshold, resulting in a more extensive search. There is evidence for the effect of consistency such that decision time increases and confidence decreases with an increase in the inconsistency of cue information (Dummel, Rummel, & Voss, 2016; Glöckner et al., 2010; Glöckner & Hodges, 2011). These effects remain even when the inconsistent information comes from less valid cues (Dummel et al., 2016; Glöckner & Betsch, 2012), suggesting that people do not ignore information from less valid cues, even in situations in which TTB would be expected to operate. Participants also searched more extensively when information incompatible with the TTB option intrudes even when specifically trained to use TTB (Söllner et al., 2014). The adaptive toolbox cannot account for the effect of information inconsistency.

In addition to evidence showing that people are sensitive to the consistency of information, there is also evidence that stopping behavior is better accounted for by a decision threshold than the stopping rules within the adaptive toolbox. TTB assumes search stops after the most valid cue that discriminates between the options, and WADD assumes exhaustive search. EAM makes predictions about stopping behavior that allow search to stop between what TTB and WADD predict, past the most valid discriminating cue but not all cues. Note, PCS does not make any predictions about stopping behavior because the formally specified version assumes all information is available. Newell and Lee (2011) showed that the amount of evidence participants required prior to stopping search matched well with predictions based on the participants' evidence accumulation threshold. Participants had lower thresholds under high costs compared to low, also matching model predictions. Hausmann and Läge (2008) also found evidence that search termination behavior was better captured by an EAM compared to a multiple strategy model. Participants tended to terminate search when the first cue's validity was over the participant's threshold but continued otherwise. In another experiment using partial information boards, where some information was initially available and the rest could be acquired, stopping behavior depended on the quality of the given information such that the probability of immediately stopping increased with increasing levels of evidence (Söllner & Bröder, 2015).

Although both single-strategy frameworks address some of the criticisms of the adaptive toolbox, they also have their own limitations. There are concerns about PCS being ill-specified. Some researchers have argued that rather than solving the strategy selection problem, PCS just replaces it with a parameter fitting problem (Marewski, 2010; Marewski & Link, 2014). This is particularly problematic because some of the processes and variables within PCS cannot be understood psychologically (Glöckner & Betsh, 2008a; Glöckner & Hodges, 2011). Moreover, the strategy selection problem is still present in the way the model addresses information search (Glöckner & Betsch, 2012; Marewski, 2010; Marewski & Link, 2014). As noted above, the fourth step of PCS requires the decision-maker to select a strategy for searching for information. This process has not been well specified. In fact, the way the PCS model has been applied in the literature assumes all information is available and does not include this fourth step (Marewski, 2010; Söllner et al., 2014). The model also fails to specify the processes involved in how the starting values for cue validity are determined, which is critical for a full account of how people make cue-based inferences.

Like the other models discussed above, EAM also has a few limitations. EAM assumes all search occurs cue-wise and so it cannot account for search that occurs alternative-wise (Newell & Lee, 2011). It’s important to note that PCS also cannot account for different search patterns because its search process is unspecified. Related to this criticism, Marewski and Link (2014) argued that EAM is limited because it has only integrated a few decision strategies and has not integrated all of the strategies in the toolbox. However, this criticism assumes all of the strategies in the toolbox are actually used, which may not be the case (Glöckner & Betsch, 2008a). Yet, one major limitation of EAM is that it does not describe the process for ordering cues for search, but instead assumes cues are ordered by cue validity. As noted above, PCS also fails to provide a process account for the initial validities of the cues used to make decisions.

2.3 Shortcomings of Current Frameworks

Concerns about the specification of the processes for determining search in both of the single-strategy frameworks discussed above are partially the result of issues with the concept of cue validity in general. All of the current frameworks rely on the concept of cue validity to explain how cues are used. The knowledge-based strategies within the adaptive toolbox assume cue use is based on validity: TTB uses validity to determine search order and WADD uses validity to weight cues. EAM also assumes search order is determined by cue validity. Even though PCS allows initial validities and final validities to differ, it still assumes weighting of the cues is based on subjective cue validity (Glöckner & Betsch, 2008a). However, the process of determining the initial subjective cue validity is not specified. Moreover, the consistency maximization process is also not specified in a way that is clear what psychological processes are involved in determining the final cue validity.

Frameworks that use cue validity to describe information use are problematic because cue validity calculations are complex, it is not clear that people actually use cue validity, nor is it clear that people can learn specific cue validities. Dougherty, Franco-Watkins, and Thomas (2008) argued that decisions based on cue validity are not psychologically plausible because cue validity calculations require that people have access to a vast amount of information in memory and require a large number of calculations. In support of these concerns, there is evidence that people are not very good at learning actual cue validities (Bergert & Nosofsky, 2007; Dieckmann & Rieskamp, 2007; Newell, Rakow, Weston, & Shanks, 2004; Newell & Shanks, 2003; Rakow, Newell, Fayers, & Hersby, 2005). In fact, in many studies of how people make cue-based inferences, participants are actually told the cue validities (Bröder, 2003; Bröder & Gaissmaier, 2007; Dieckmann & Rieskamp, 2007; Dummel et al., 2016; Glöckner & Betsch, 2012; Hausmann & Lage, 2008; Newell & Shanks, 2003; Rieskamp & Hoffrage, 2008; Söllner et al., 2014). In studies where participants are not told cue validities, there is evidence that validity does not provide the best account of cue use, instead participants' behavior is best accounted for by a combination of validity and discrimination rate (Rakow et al., 2005; Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003). Because of issues with cue validity, Dougherty, Franco-Watkins, and Thomas (2008) argued that cue use is likely driven by memory retrieval.

Moreover, there is evidence for the role of memory in cue-based inferences. A number of studies have shown that memory accessibility influences decision making through apparent strategy selection. There is evidence for different strategies when accessibility is directly manipulated. In studies in which accessibility or salience of information was manipulated, participants' decisions matched compensatory strategies when less valid cues were more accessible or salient and matched noncompensatory strategies when more valid cues were more accessible or salient (Platzer & Bröder, 2012; Platzer et al., 2014). Participants' decisions were also found to match compensatory strategies when the frequency with which they saw cues during training was either negatively correlated or uncorrelated with cue validity (Lawrence et al., 2018a). Overall, studies suggest that when less valid cues are more accessible, people are more likely to use compensatory strategies over noncompensatory ones, providing evidence for the importance of accessibility in theories of cue-based inferences.

Evidence for the influence of accessibility on cue preferences also supports the claim that accessibility is important in cue-based inferences. Although validity and accessibility are typically confounded in studies of cue-based inferences, there have been recent attempts to directly manipulate accessibility. Lawrence et al. (2018a) manipulated the relationship between cue validity and the frequency of cue presentations during training. Participants selected the two most valid cues more often when frequency and validity were positively correlated compared to when they were negatively correlated. Participants also selected the two least valid cues more when frequency and validity were negatively correlated compared to when they were positively correlated. Another study also found the serial position of correct discriminations across training blocks affected cue use (Lawrence et al., 2018b). This study tested serial position effects (primacy/recency) by having pairs of cues with the same validity but one member correctly discriminated more at either the beginning (primacy) or the end (recency) of each training block and the other member correctly discriminated more in the middle of each block. Participants selected the cue that correctly discriminated more at the beginning compared to the one that correctly discriminated more in the middle, despite the cues having the same validity. Again, this provides evidence for an accessibility effect for cue selection. In general, studies that directly manipulated accessibility while controlling for validity have found an effect of accessibility on cue preference and use that cannot be accounted for within the existing frameworks.

CHAPTER 3: ALTERNATIVE ACCESSIBILITY-BASED FRAMEWORK

An alternative framework based on memory accessibility likely provides a better account of how people make cue-based inferences. Rather than assuming cue use is based on a precomputed cue hierarchy, the accessibility-based framework assumes cue use is based on memory retrieval. Accessibility can be influenced by memory phenomenon, such as repetition effects, but it can also be influenced by cue validity. In this way, the framework can account for the effects of accessibility on cue use found in previous studies (Lawrence et al., 2018a, 2018b; Platzer et al., 2014) that current frameworks cannot explain. The accessibility-based framework should also be able to account for apparently different strategies via differences in which cues are accessible at the time of the inference.

The accessibility-based framework for cue-based inferences is an extension of the HyGene model (Thomas et al., 2008). The original HyGene was developed as a process model of hypothesis generation (Thomas et al., 2008) and has been extended to cue-based inferences (Lawrence et al., 2018a). Briefly, the original model is based on three basic principles. First, data in the environment function as retrieval cues that prompt retrieval of hypotheses from long-term memory. Second, the number of hypotheses that can be considered is limited by cognitive constraints and task characteristics. Third, these hypotheses are then used to judge probability and drive information search.

Specifically, the original version of HyGene is based on three memory constructs: working memory, exemplar memory, and semantic memory. These memory constructs interact in the process of generating hypotheses. First, a prototype hypothesis is extracted based on a comparison of information in the environment to exemplars in memory. Exemplar memory is made up of imperfectly encoded representations (traces) of a decision maker's past experiences with observations and hypotheses. These traces are activated based on the similarity between them and the information currently in the environment. Those traces that are activated above a threshold are amalgamated to create an unspecified probe based on the hypotheses in exemplar memory most similar to the data observed. Then the unspecified probe is compared to representations in semantic memory, generalized knowledge representations of data and hypotheses. Those hypotheses in semantic memory that exceed a similarity threshold when compared to the probe are maintained in working memory, but the number of hypotheses in working memory is limited. Finally, the probability of each hypothesis in working memory is determined by its activation relative to the activation of all hypotheses in working memory.

Recently, HyGene has been extended to cue-based inferences (Lawrence et al., 2018a). The basic principles and processes described above still apply but rather than generating hypotheses, the model generates cues for cue-based inferences. The three principles can be extended to cue-based inferences. First, data in the environment function as retrieval cues that prompt retrieval of cues from long-term memory. Second, the number of cues that can be considered is limited by cognitive constraints and task characteristics. Third, the cues that are actively being considered are used as input into a decision strategy.

The specific processes described for the original version of HyGene also operates here. First, information in the environment (e.g., the goal of task and the objects) prompt the retrieval of cues. Specifically, cues are retrieved based on a comparison between information in environment and exemplar memory. Information in the environment can be characterized as the goal of choosing the option with the higher criterion value; however, other goals could also be instantiated. Exemplar memory is made up of previously experienced decisions with those cues. Each previously experienced object is encoded individually with either a win or loss feedback signal depending on if that object was higher or lower on the criterion than whichever object it was compared to. The quality of the encoded information in exemplar memory is based on an encoding parameter (*L*) such that a value of 1 means the information has been perfectly encoded and a value of 0 means that the information is completely degraded. This allows the model to account for individual differences at the level of learning.

The comparison of information in the environment (goal context) and exemplar memory results in the creation of an unspecified probe. In this model, unspecified probes are actually created for both win and loss contexts, these can be thought of as the prototypical winning (or losing) option. Then the activation of individual cues (A_s) is calculated based on the similarity between the semantic representation of the cues and each unspecified probe. Overall cue activation is calculated by taking the difference in the memory activation between win and loss contexts. Thus, cues with a higher activation in the win context relative to the loss context are more accessible overall. Importantly, the memory representation of the learning environment assumes that people are sensitive to other factors in the environment that influence accessibility in addition to cue validity. For example, repetition effects can be accounted for in the way the framework structures exemplar memory (Lawrence et al, 2018a), more exemplars with that cue results in that cue being more accessible. Yet, the framework is still sensitive to the accuracy of the cues through the way activation is calculated in general.

Then cues are generated into working memory one at a time based on activation (A_s) with only cues that exceed the activation threshold (Act_{minH}) being generated. The order in which cues are generated into working memory is probabilistic, with the probability determined by the relative activation of the cue to all other cues. This means that more accessible cues are likely generated early, but the order is not deterministic. The activation threshold (Act_{minH}) is updated throughout the generation process to match the activation of the cue in working memory with the highest activation. Cues stop being generated into working memory when the number of retrieval failures exceeds the threshold or there is no time left, representing memory and task constraints. Then those cues that have been generated into working memory are used in inference decisions.

The accessibility framework presented in this dissertation is nearly the same as the version of HyGene described in the Lawrence et al. (2018a), but with one important difference. The previous version assumed that cues could be fed into a number of qualitatively different strategies, such as TTB or WADD. However, in this dissertation, a single strategy is assumed because of issues with the multi-strategy approach described above (i.e. complexity and difficulty falsifying). Instead, the accessibility-based framework assumes that cues are integrated in a compensatory manner with the cues weighted by accessibility. This assumption is based on a number of studies that suggest compensatory strategies may be the default strategy (Bröder, 2003; Glöckner & Betsch, 2008a; Rieskamp & Otto, 2006; Söllner et al., 2014). However, which cues are actually used in the compensatory decision depends on which cues are available in working memory. This means that the model can show adaptive behavior and mimic apparently different strategies based on which cues are available in memory.

The effects found to result in qualitatively different strategies in the adaptive toolbox literature can be accounted for with the accessibility framework. Evidence that people are more likely to use noncompensatory strategies under high information cost (Bröder, 2000; Bröder & Gaissmaier, 2007; Bröder & Schiffer, 2003; Newell & Shanks, 2003) and under time pressure (Glöckner & Betsch, 2008b; Rieskamp & Hoffrage, 2008) can be accounted for by the threshold for the number of retrieval failures before cue generation stops. This results in fewer cues being available to make a decision, resulting in decisions that look like noncompensatory ones. Effects of cue dispersion, more dispersed cue validities leading to more noncompensatory decisions (Bröder, 2003; Newell & Shanks, 2003), can be accounted for by the minimum activations of cues. Cues that are more accessible than others are more likely to enter working memory first and keep less accessible cues out, resulting in more noncompensatory type of behavior. In general, adaptation to the ecological structure (Bröder, 2003; Lawrence et al., 2018a; Rieskamp & Hoffrage, 2008; Rieskamp & Otto, 2006) is the result of changes in the accessibility of cues based on previous experiences with the cues.

Like the single-strategy frameworks discussed above, the accessibility framework addresses the criticisms of the adaptive toolbox: eliminating the need for strategy selection, providing a falsifiable model, and being able to account for individual differences. However, the accessibility framework addresses these criticisms better than the other frameworks. The parameters in the accessibility framework have psychological grounding, eliminating the criticism that parameter fitting replaces strategy selection that has been levied against PCS (Marewski, 2010; Marewski & Link, 2014). Further, the framework makes predictions that are falsifiable because the predictions are based on well-established memory research and the way the model is typically instantiated results in one free parameter (*L*). The accessibility framework can also demonstrate individual differences through the encoding parameter. Unlike EAM or PCS, the accessibility framework can also account for differences in search order because search order is not deterministic in the accessibility framework.

The accessibility framework has several other similarities with the single-strategy models while providing a better process account of cue-based inferences. Like the evidence accumulation model (Lee & Cummins, 2004), the accessibility framework allows for decisions based on any number of cues, which matches empirical evidence that people often do not stop searching at the first discriminating cue nor do they always search all cues (Newell & Shanks, 2003; Newell et al., 2003). Yet, the search order and processes involved in search are better specified in the accessibility framework because it is not based on precomputed validity. Similar to PCS, the accessibility framework is based on prior experience (Betsch & Glöckner, 2010) and assumes decisions are based on all the evidence available to the decision maker (Betsch & Glöckner, 2010; Glöckner & Betsch, 2008a). However, unlike PCS, the processes accounting for the influence of prior experience and the determination of which cues are available are well specified in the memory representation of the learning environment.

Overall, the accessibility framework provides a better theoretical account of cue-based inferences than the current frameworks. In addition to addressing issues with the adaptive toolbox, it also addresses criticisms of the current single-strategy frameworks (e.g. poor specification of search, too many parameters). Unlike the other frameworks, the accessibility framework is not based on a precomputed cue hierarchy but is instead based on memory accessible. Moreover, the process for cue generation is well specified and psychologically plausible because it is based on well-established memory phenomenon. Thus, the model can account for both the influence of direct and indirect memory manipulations as well as accounting for adaptive decision-making behavior.

CHAPTER 4: SIMULATION OF ACCESSIBILITY-BASED FRAMEWORK

Before getting into the specifics of the experiments testing the accessibility framework, a simulation will be described. The goal of this simulation was to show that the accessibility framework mimics apparently different strategies without explicitly specifying qualitatively different strategies. Specifically, the simulation shows that the model behaves differently under highly dispersed cue validities compared to cue validities that are closer together. As noted above, cue validity dispersion is an environmental factor that has led to different strategy classifications with more dispersed cue validities leading to more noncompensatory classifications (Bröder, 2003; Newell & Shanks, 2003). The extension of the HyGene model described above was used to simulate the behavior of the accessibility framework. With this implementation of HyGene, the simulation provides predictions in terms of cue preferences, the number of cues searched, and decision behavior.

4.1 Simulation Method

Three data sets were simulated to use as training data for the HyGene model. These data sets included 200 pairs of options with each option having values for five dichotomous cues. The data sets differed in the dispersion of cue validities but had the same average cue validity as shown in Table 1. All cues had a discrimination rate of .6, meaning that for each cue the value differed between the options on 60% of the trials. The model was trained on each dataset using the method described in the model section.

Each option in the training dyads was encoded with either a loss context or win context, depending on which option had a higher criterion value. These traces were used to determine the overall activation for each cue. Then the normalized activations were used to give the probability of each cue being retrieved at the time of the decision.

Table 1
Validity of Cues for High and Low Dispersion Data Sets

	Cue 1	**Cue 2**	**Cue 3**	**Cue 4**	**Cue 5**
High	.91	.78	.70	.62	.51
Med	.77	.73	.70	.67	.63
Low	.72	.71	.70	.69	.68

Fifty test trials were also simulated to look at the decision behavior of the model. These were presented as dyads in which the model must select which option is higher on the criterion based on the cue values. Importantly, the activations from the training data were used to determine which cues were available to the model at the time of the decision. The cues that were available in working memory can be thought of as the cues searched by the model. Because of the way the generation process operates, a varying number of cues may be used by the model for each decision. Only cues that have been generated into working memory were used in the decision process in which cue values were weighted by their normalized activations and summed. The option with the highest total was selected. If the options were tied, then a selection was made randomly.

In addition to manipulating the dispersion of validity in the training data, the encoding parameter and the threshold for the number of retrieval failures were also manipulated within the model. The encoding parameter was manipulated to show the differences in model performance under different levels of learning. The threshold for the number of retrieval failures was manipulated to demonstrate the effect of time constraints on the model.

The way the model operates means that it should mimic both noncompensatory and compensatory behavior. When one cue is much more valid than other cues, as is the case for the high dispersion dataset, its probability of being generated into working memory first is high. Once this best cue is generated, other cues cannot exceed the activation of that cue so the best cue will be the only one active in working memory. If that cue fails to discriminate between the options, then the model randomly selects one of the options. In contrast, when cues are much closer in validity, the best cue may not be generated first. This results in more cues entering working memory and decisions that are more likely to match a compensatory strategy. Thus, the simulations should show differences depending on cue validity dispersion: less valid cues are used more frequently, more cues are used, and decisions match a compensatory strategy more often in the low dispersion dataset compared to the high dispersion data set. Establishing that the model can mimic different decision strategies is important for justifying it as an alternative to the current frameworks.

4.2 Simulation Results

In general, the number of cues generated increased as the dispersion of validity decreased (as shown in Table 2). This pattern was stronger for higher levels of the encoding parameter and when the threshold for the number of retrieval failures was higher. Importantly, this matches the pattern that would be expected from different decision strategies without invoking qualitatively different strategies. The number of cues generated was also higher when the threshold of retrieval failures was higher. Again, this matches the idea that under time pressure people are more likely to use fewer cues. Finally, as the encoding parameter decreased, the number of cues selected tended to decrease. The differences in the number of cues generated for the levels of dispersion of cue validity also decreased as the encoding parameter decreased.

Table 2

Average Number of Cues Select by Data Set, Threshold for Retrieval Failures, and Encoding Parameter

	Threshold = 3				
	L=1	L=.8	L=.6	L=.4	L=.2
High	1.64	1.65	1.63	1.64	1.61
Med	1.78	1.75	1.73	1.68	1.61
Low	1.78	1.78	1.74	1.68	1.62

Table 2 (continued).

	Threshold = 10				
	L=1	L=.8	L=.6	L=.4	L=.2
High	1.73	1.75	1.74	1.72	1.71
Med	1.96	1.93	1.87	1.82	1.70
Low	2.01	1.98	1.90	1.82	1.70

In addition to looking at the average number of cues generated, the different model conditions were also compared in terms of the proportion of decisions based on one cue, a sign of noncompensatory behavior. As the dispersion of validity increased, the number of decisions based on one cue also increased as shown in Table 3. This is in line with evidence for the use of more noncompensatory strategies when the dispersion is high. When the threshold for the number of retrieval failures was lower, the proportion of decisions based on a single cue was higher. Again, this matches the idea that under time pressure people are more likely to use fewer cues. Encoding parameter and the dispersion of validity appeared to interact, such that the encoding parameter did not have much of an effect on the proportion of times a single cue was generated when dispersion was high. However, when dispersion was low or at a medium level, the proportion of decisions based on a single cue increased as the encoding parameter decreased. At the lowest encoding parameter, the three dispersion conditions were nearly equal.

Table 3

Proportion of Decision Based on a Single Cue by Data Set, Threshold for Retrieval Failures, and Encoding Parameter

	Threshold = 3				
	L=1	L=.8	L=.6	L=.4	L=.2
High	0.48	0.47	0.48	0.48	0.50
Med	0.41	0.42	0.43	0.46	0.49
Low	0.41	0.41	0.43	0.46	0.49

	Threshold = 10				
	L=1	L=.8	L=.6	L=.4	L=.2
High	0.42	0.41	0.41	0.42	0.43
Med	0.31	0.33	0.35	0.38	0.44
Low	0.30	0.31	0.34	0.38	0.44

In terms of cue preferences, preference for the most valid cue was highest in the high dispersion condition and lowest in the low dispersion condition (see Figure 1). Conversely, preference for the lowest validity cue was strongest in the low dispersion condition and weakest in the high dispersion condition. Again, showing that the model can mimic different decision strategies. When the threshold for the number of retrieval failures was higher, preference for the most valid cue was generally higher. However, the threshold did not have a strong effect on the overall pattern of cue preferences. This is consistent with the idea that people simply select more cues when they have more time. As the encoding parameter decreased, preference for the more valid cues decreased while preference of the least valid cues increased. Differences in preferences between the different dispersion conditions also tended to decrease.

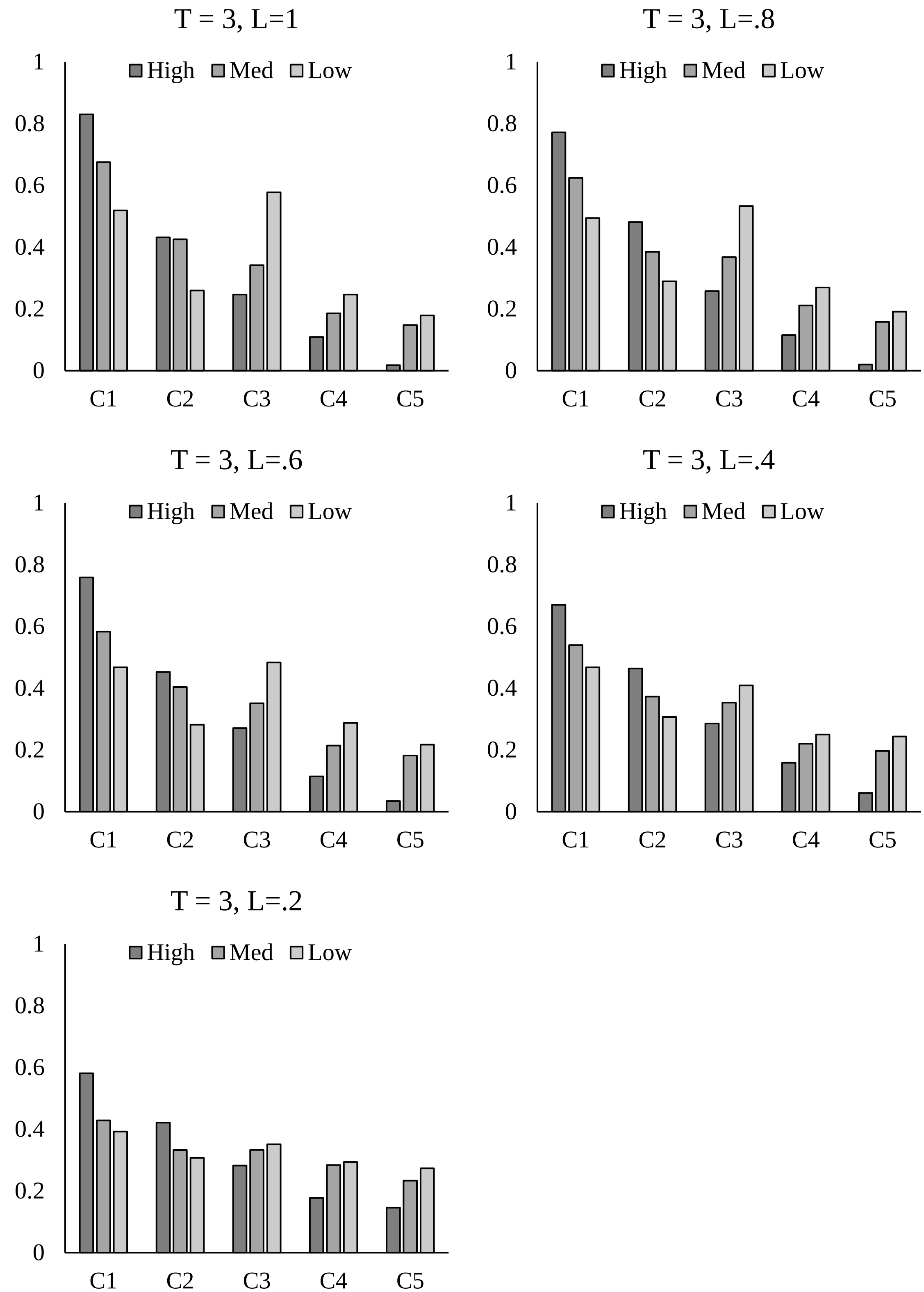

Figure 1. Proportion of trials each cue was selected by data sets and encoding parameter.

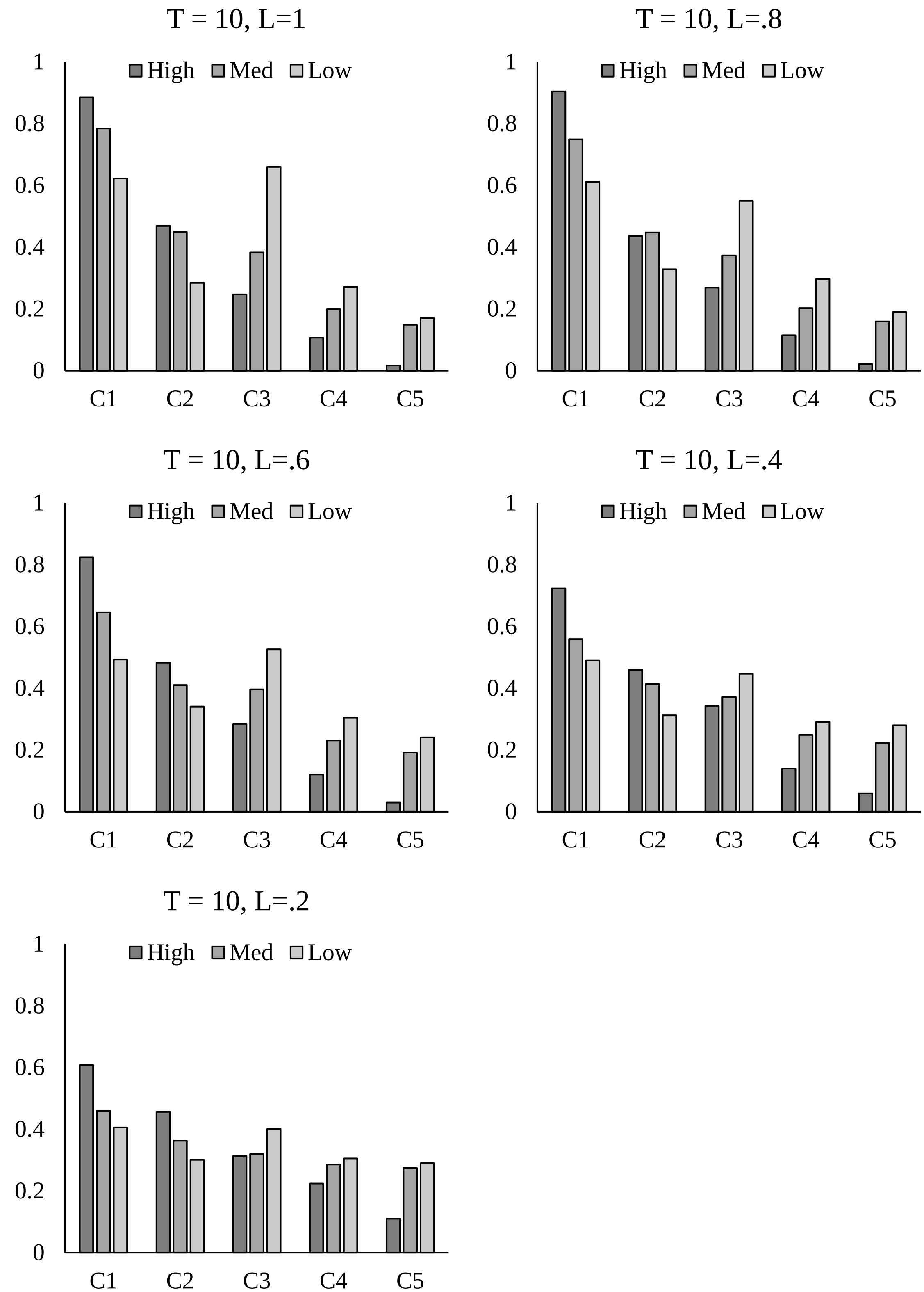

Figure 1continued.

For decision outcome, as the dispersion of validity decreased, the proportion of decisions matching a compensatory strategy increased (Table 4). Again, showing that the model can mimic different decision strategies without actually invoking qualitatively different strategies. The manipulation of the threshold for the number of retrieval failures did not have a large effect of apparent decision strategy. In general, as the encoding parameter decreased, the proportion of decisions matching a compensatory strategy increased. This is likely because differences in preferences between the different cues generally decreased as the encoding parameter decreased, resulting in fewer decisions based on the most valid cues.

Table 4
Proportion of Decisions Matching Compensatory Rule When Compensatory and Noncompensatory Differ by Data Set, Threshold of Retrieval Failures, and Encoding Parameter

Threshold = 3					
	L=1	L=.8	L=.6	L=.4	L=.2
High	.27	.31	.32	.39	.46
Med	.40	.44	.48	.50	.58
Low	.52	.54	.57	.56	.60

Threshold = 10					
	L=1	L=.8	L=.6	L=.4	L=.2
High	.25	.25	.30	.37	.47
Med	.35	.39	.47	.53	.59
Low	.51	.51	.58	.58	.61

4.3 Simulation Discussion

The above simulations demonstrate that the accessibility-based framework mimics changes in apparent decision strategy without invoking qualitatively different strategies. The model demonstrated differences in cue preferences, the number of cues generated, and option selected under training sets with different dispersions of validity. In general, the model showed more compensatory type behavior when dispersion was low: less preference for the most valid cue, more cues selected, and more decisions matching a compensatory type of decision. The model also showed shifts in apparent decision strategy toward compensatory types of decisions when the maximum number of retrieval failures decreased, a proxy for time pressure. Overall, the model demonstrates many of the common findings in the cue-based inferences literature, such as more decisions that seem compensatory under time pressure (Bröder & Gaissmaier, 2007; Bröder & Schiffer, 2003; Rieskamp & Hoffrage, 2007) and under less dispersed validity conditions (Bröder, 2003; Newell & Shanks, 2003). These simulations essentially provide a proof-of-concept for the experiments that follow.

CHAPTER 5: EXPERIMENTS

Three experiments were conducted to test whether the accessibility framework provides the best account of adaptive decision making in cue-based inferences. As mentioned above, a few experiments have already shown that accessibility affects how cues are used (Lawrence et al., 2018a, 2018b; Platzer et al., 2014). However, these experiments fail to demonstrate that a memory-based framework truly provides the best account in terms of both information search and decision. Prior research has focused on cue preference with only a little attention paid to the decision processes. Further, research specifically looking at decisions have been framed within the multiple-strategy framework, arguing that accessibility affects strategy selection. The overall goal of the following experiments was to show that the accessibility framework accounts for differences in decisions processes that result from both direct manipulations of cue accessibility and manipulations of the decision environment that commonly result in different apparent strategies.

5.1 Experiment 1: Serial Position Effects in Cue-Based Inferences

The goal of the first study in this dissertation was to show that the accessibility framework can account for apparent strategy differences that result from both a direct memory manipulation and the manipulation of cue validity. In this study, memory was manipulated via the serial position of the correct discriminations of cues, which is when a cue correctly indicates which option is higher on the criterion. The manipulation of serial position is based on evidence in the memory literature that the serial position of learned items affects later recall of that information (Bjork & Whitten, 1974). Mean cue validity and dispersion of cue validities were also manipulated based on evidence that these affect strategy selection (Bröder, 2003; Newell & Shanks, 2003). It was hypothesized that the effects of both manipulations would match predictions based on the accessibility-based framework. Differences in cue preferences and final decision that result from these manipulations should be caused by differences in the accessibility of the cues in memory. The other frameworks discussed in this proposal have no mechanism for accounting for the influence of a direct memory manipulation on cue use and decision in cue-based inferences.

Long-term serial position effects have been well established in the memory literature. When learning word lists, people typically show both long-term primacy and recency (Bjork & Whitten, 1974). Specifically, when presented with a number of word lists, people tend to recall words presented at the beginning (primacy) and ends (recency) of each list better than information in the middle of each list. Both primacy and recency have also been demonstrated to influence decision-making (Hastie & Park, 1986; Hogarth & Einhorn, 1992; Peterson & DuCharme, 1967) and hypothesis generation (Lange, Thomas, Buttaccio, Illingworth, & Davelaar, 2013; Lange, Thomas, & Davelaar, 2012).

Prior research has also shown that primacy affects cue preferences in cue-based inferences (Lawrence et al., 2018b). In that study, the serial position of correct cue discriminations was manipulated within training blocks. During training, participants were asked to choose which company had a higher stock based on five cues for each company. The position of correct discriminations was manipulated for two pairs of cues.

Cues were paired such that they had approximately the same validity, but one member of the pair correctly discriminated more often in the middle of each training block and the other correctly discriminated more in the beginning (primacy) or end (recency). The two experiments conducted in that study showed that people selected the cue that discriminated more frequently at the beginning compared to the one that discriminated more frequently in the middle, even when the middle cue had the same or slightly higher validity than the primacy cue. Although the exact mechanisms for this effect were not directly tested, the authors concluded that this may have been due to an encoding advantage for cues learned early in each block.

Yet, the two experiments reported in that paper found slightly different results; the primacy effect was stronger in Experiment 2 compared to Experiment 1. The authors speculated that this difference was due to the cue validities being less dispersed in Experiment 2 compared to Experiment 1. Specifically, they argue that when the cue validities are greatly dispersed, people might not show a strong preference among the remaining cues once the two most valid cues were selected. As noted above, there is evidence that people use apparently different strategies when cues have more dispersed validities, specifically using more noncompensatory strategies (Bröder, 2003; Newell & Shanks, 2003). However, the previous study did not systematically manipulate cue validity. Moreover, this study was focused on cue preferences and did not make any hypotheses about the effect of the manipulations on final decisions.

The current study provides a more systematic test of the influence of serial position and the dispersion and mean of cue validities in cue-based inferences than the previous study. Specifically, the serial position manipulation was implemented using a similar method as Lawrence et al. (2018b). However, there were two serial position conditions: one in which the more valid cue pair included the primacy cue and the less valid cue pair included the recency cue and one in which the less valid cue pair included the primacy cue and the more valid cue pair included the recency cue. There were also two conditions for mean cue validity with the average validity of the cues being either high or low. Finally, there were two conditions for cue dispersion, where the cue validities were either highly dispersed or not. All of these manipulations were crossed resulting in eight experimental conditions. By systematically manipulating these variables, the experiment provides a more complete test of the accessibility framework by testing whether it correctly predicts the effects of both direct manipulations of memory and manipulations of the cue validity structure as well as the interaction between these variables. Further, this experiment tested the effect of these manipulations, not only on cue preferences, but also on apparent decision strategy by looking at decisions between companies at test.

The accessibility framework makes several predictions in regards to both cue preference and actual decision. First, the manipulations were predicted to affect cue preference by affecting the accessibility of the cues. Recall from the description of the framework that cue accessibility is based on both the accuracy of the cues and general memorial effects. Thus, the serial position of correct cue discriminations should affect cue preference, such that participants generally prefer the primacy cue over the similarly valid middle cue as found in previous research (Lawrence et al., 2018b). A recency effect is not predicted based on previous research that has failed to find such an effect but was manipulated in case the effect interacted with validity manipulations. There are no specific predictions about the effect of whether the primacy cue is in the more valid cue pair or the less valid. Yet, the strength of the effect of primacy is predicted to interact with the manipulations of cue validity. The effect should be stronger when all cues have more similar validities. This may be especially pronounced when the mean validity is lower. The reason for these predicted interactions is that the primacy effect may be unable to provide a noticeable boost to the accessibility in conditions in which the best cues are highly valid.

The accessibility of the cues should also affect participants' decisions, leading to apparently different strategies. The dispersion of the cue validities should affect apparent strategy such that more dispersed cue validities should result in more decisions matching a noncompensatory type strategy; this may be more pronounced when the average validity is higher. Conversely, less dispersed cue validities should result in more compensatory type of behavior, which may not be influenced by average cue validity. This prediction is based on the functioning of the accessibility framework such that when cues are much more valid than other cues, only the most valid cue is available at the time of the decision. The serial position manipulation should also affect apparent strategy such that when the more valid cue pair includes the primacy cue, decisions should become more noncompensatory compared to when the less valid cue pair includes the primacy cue. Again, this is because the primacy manipulation should provide a boost to the accessibility of a cue that is already relatively more accessible than the other cues because of its validity, making it more likely to be the only cue available. A boost to a less valid cue makes it more accessible in general but may not make it so much more accessible than other cues that it is the only one available.

5.1.1 Method

There were 308 participants (approximately 39 participants in each condition) recruited for this study using the online experiment management system at Georgia Institute of Technology. Participants received course credit for their participation and could receive $5.00 for correctly selecting the company with the highest stock price on one randomly selected trial from the test phase.

The design of this experiment was a 2 (serial position) by 2 (mean cue validity) by 2 (dispersion of cue validity) between subjects design. The two conditions for the serial position manipulation were either the primacy cue was in the most valid cue pair and the recency cue was in the less valid cue pair or vice versa. The two conditions for the mean cue validity were .8 (high) and .65 (low). The two conditions for the dispersion of cue validity were a standard deviation of .05 (low) or .12 (high). This resulted in 8 experimental conditions in the experiment.

Participants completed both a training phase and a test phase of a stock-forecasting task in which they were asked to predict which of two companies would have a higher stock price based on the attributes of the companies. The training phase consisted of 5 blocks of 15 dyad comparisons. The test phase consisted of 40 dyad comparisons. There were five dichotomous attributes (cues) that were associated with each company: asset rating (AR), earning potential (EP), liquidity appraisal (LA), optimized capital (OC), and profit intensification (PI). The validities of the cues were manipulated such that there were two pairs of cues with similar validities and one cue with a unique validity. The mean and dispersion of the cue validities were manipulated, as shown in Table 5, resulting in four different validity conditions.

Table 5
Validity of the Cues by Condition for Experiment 1

	High Mean/ Low SD	High Mean/ High SD	Low Mean/ Low SD	Low Mean/ High SD
Cue 1	0.84	0.96	0.71	0.8
Cue 2	0.84	0.96	0.71	0.8
Cue 3	0.8	0.76	0.64	0.58
Cue 4	0.8	0.76	0.64	0.58
Cue 5	0.71	0.6	0.56	0.51
Mean	.80	.81	.65	.65
SD	.05	.13	.06	.12

The frequency with which each cue correctly discriminated in the beginning (first five trials 1-5), middle (trials 6-10), and end (last five trials 11-15) of each block of the training phase were also manipulated. One member of each pair correctly discriminated more often in the middle of every block than either the beginning or the end (Cue 1 and Cue 3). The other member of the pair correctly discriminated more at the end or the beginning of each block as shown in Table 6. The cue that was the primacy cue (more discriminations early) and the cue that was the recency cue (more discriminations at the end) were manipulated so that in one condition the primacy cue was in the higher validity pair and in the other condition the primacy cue was in the lower validity pair. This was done to isolate the effects of the serial position of cue discrimination on cue use while controlling cue validity. Note that Cue 5 was a filler cue that was not specifically manipulated.

Table 6
Average Number of Discriminations by Position in Each Training Block

High Mean/ Low SD	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5
Begin (1-5)	2.4	4.8	2.2	1	2
Middle (6-10)	2.8	1.8	3	1.8	2.4
End (11-15)	2.4	1	2	4.4	2

High Mean/ High SD	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5
Begin (1-5)	2.6	5	2.2	1	1.2
Middle (6-10)	3.2	2	2.6	1.8	2.2
End (11-15)	2.8	1.6	2	4	2

Table 6 (continued)

Low Mean/ Low SD	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5
Begin (1-5)	1.8	4	2	0.8	1.4
Middle (6-10)	2.8	1.4	2	1	2.4
End (11-15)	1.8	1	1.8	4	1.4

Low Mean/ High SD	Cue 1	Cue 2	Cue 3	Cue 4	Cue 5
Begin (1-5)	1.8	4	2	1	1
Middle (6-10)	3.2	2	2	1	2.2
End (11-15)	2.2	1.2	1.2	3.2	1.4

Note: Order manipulation flips the beginning and ending trials in each block.

To further explicate the method, note that Cue 1 and Cue 2 had equivalent validities, but Cue 2 had more correct discriminations at the beginning of each block compared to Cue 1 (in the condition in which the primacy cue is in the most valid cue pair). For Cue 3 and Cue 4, note that both had equivalent validities, but that Cue 4 had more correct discriminations at the end of each training block compared to Cue 3. If cue usage is dependent on long-term serial position effects, then participants should choose Cue 2 over Cue 1 (primacy) and may choose Cue 4 over Cue 3 (recency) in this condition. Overall discrimination rate (combining both correct and incorrect discriminations) was held constant for Cues 1 through 4, meaning the values of individual cues differed between companies for about 60% of the trials.

During the training phase, participants were presented with 75 company dyads presented in 5 blocks. Each block was labeled as corresponding to a specific market sector: technology, financial, utility, property, or healthcare. There were no other differences between blocks other than the label. The labels and the spatial locations of the cues were random between participants, meaning the cues listed in Table 6 did not appear in the same location or have the same label for all participants. The participants were asked to predict which of the two companies presented would have a higher stock value based on the attributes, as shown in Figure 2. Once they selected which company they thought would have a higher stock price they were given feedback on whether they were correct.

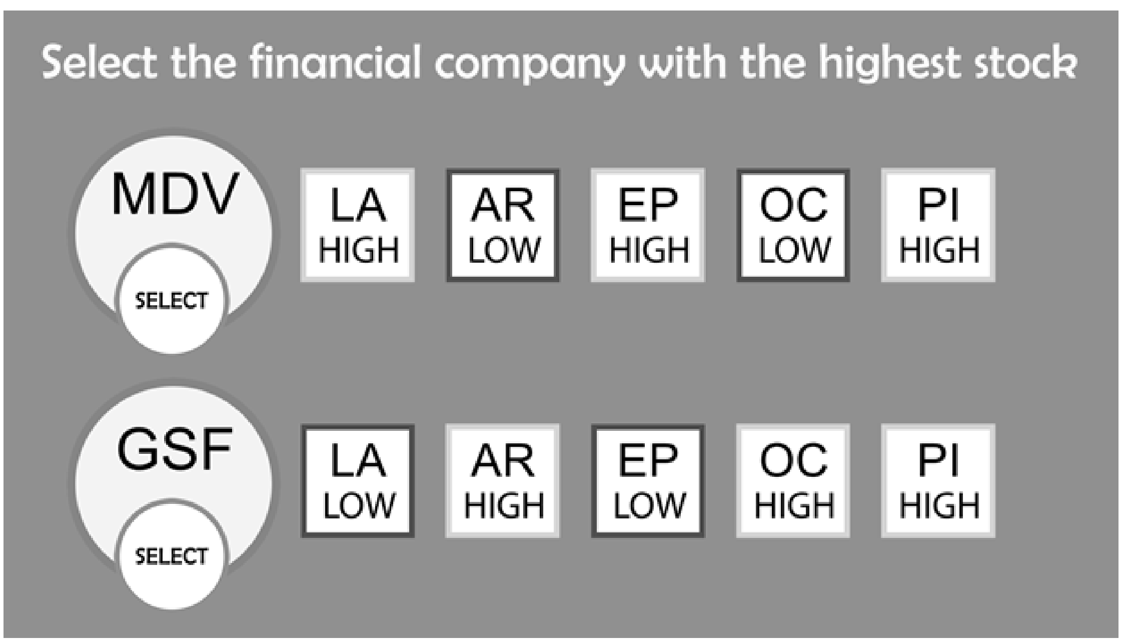

Figure 2. Example of a single learning trial in Experiment 1.

Between each of the blocks, participants completed a 25-second distractor task in which they were asked to verify simple mathematical equations involving addition or subtraction of single digit numbers. Each equation was presented for 5 seconds and the participants indicated whether the solution was correct.

In the test phase, participants were presented with 40 dyads and asked to indicate which of the two companies they thought would have the higher stock value, like the training phase but without feedback. These companies were labeled as belonging to the service market sector, a sector not seen during training. Before comparing each company dyad, participants were told to select which cues they would like to use based on what they learned in the previous sectors. The number of cues that they had to select before they compared dyads varied between 1 and 5. The number of cues participants had to select was manipulated to determine participant cue preference. The participants were given 8 test trials (dyads) for each required number of cues. For the cases in which the number of cues was less than 5, the participants clicked on the cue labels to select the cues that would be revealed on the next screen during their choice between the dyads. On the 8 trials in which they saw all the cues, 5 of those were structured to discriminate between compensatory and noncompensatory decisions.

5.1.2 Results

5.1.2.1 Learning

Learning was checked using a generalized linear model with a binomial logit link to test whether block, order condition, dispersion of validity, and mean validity affected performance during the training trials. Block (χ^2 (4) = 36.90, p < .0001) significantly affected performance during training as shown in Figure 3. Planned comparisons for the effect of block showed that the odds of selecting the correct company were higher in the last block (*O* = 1.80) compared to the first block (*O* = 1.47, *p* < .001), suggesting that participants were able to learn.

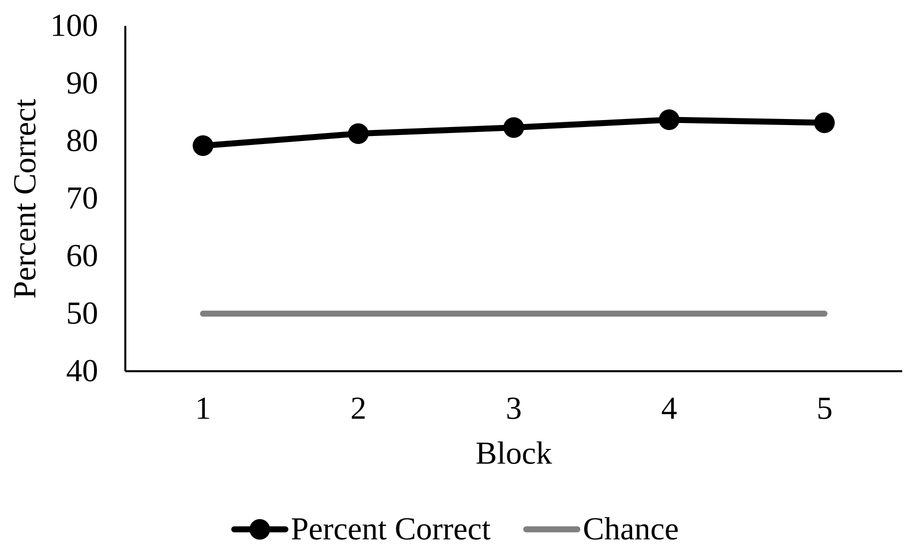

Figure 3. Percent correct during training across blocks for Experiment 1.

Order condition (χ^2 (1) = 5.67, p = .017), mean validity (χ^2 (1) = 262.80, p < .0001), and dispersion of validity (χ^2 (1) = 42.33, p < .0001) also significantly affected performance during training such that the odds of being correct were higher in the group in which primacy was manipulated in the less valid cue pair (*O* = 1.74, 82.60%) compared to when it was in the most valid cue pair (*O* = 1.66, 81.32% correct), in the high mean group (*O* = 2.44, 91.71% correct) compared to the low mean group (*O* = .96, 72.24% correct), and in the high dispersion group (*O* = 1.84, 83.10% correct) compared to the low dispersion group (*O* = 1.57, 80.85% correct). There was also a significant interaction between dispersion of validity and mean validity (χ^2 (1) = 36.10, p < .001), as shown in Figure 4. The dispersion of validities appeared to have a stronger influence on performance when the mean validity was high. There was also a significant interaction between mean validity and block (χ^2 (4) = 9.52, p = .049) such that those in the high validity condition improved more across early blocks than those in the low validity condition. The differences in performance at training between the different mean conditions reflect the fact that the decision environment was more challenging in the low mean condition and does not necessarily indicate that those in the low mean condition could not learn. The other group differences at training were small and all conditions showed training performance above chance, suggesting that participants were able to learn.

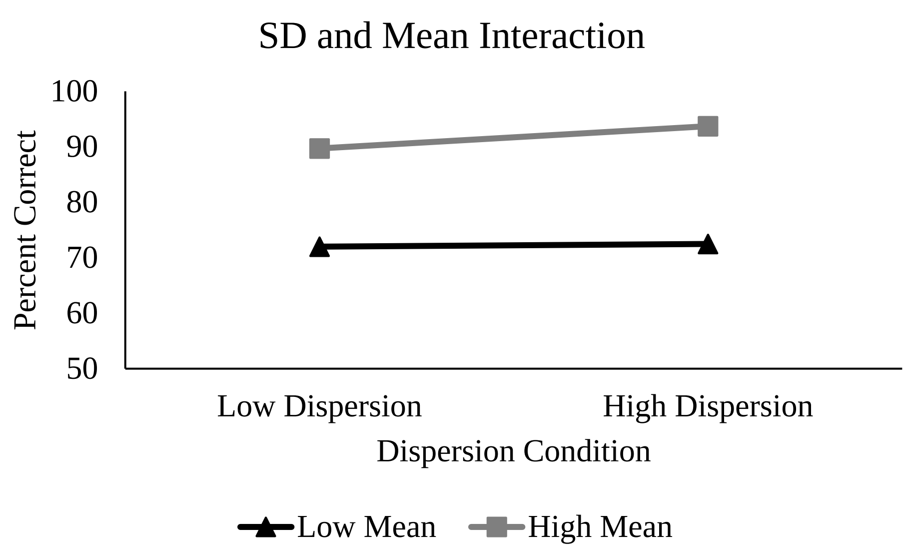

Figure 4. Percent correct during training by dispersion condition and validity condition.

5.1.2.2 Cue Preference

Figure 5 shows cue preference for each cue by condition. A generalized linear model with a binomial logit link was used to test the effect of order condition, dispersion of validity and mean validity on cue preferences. The number of cues participant had to choose was also included as a predictor. Specifically, the model was predicting preference for the primacy cue (or recency cue) over its similar validity middle cue. Thus, it was only run on trials in which participants selected either the primacy cue or its similar validity middle cue, excluding cases in which none of the cues were selected or both of the cues were selected. Separate analyses were conducted to test preference for the primacy cue and preference for the recency cue.

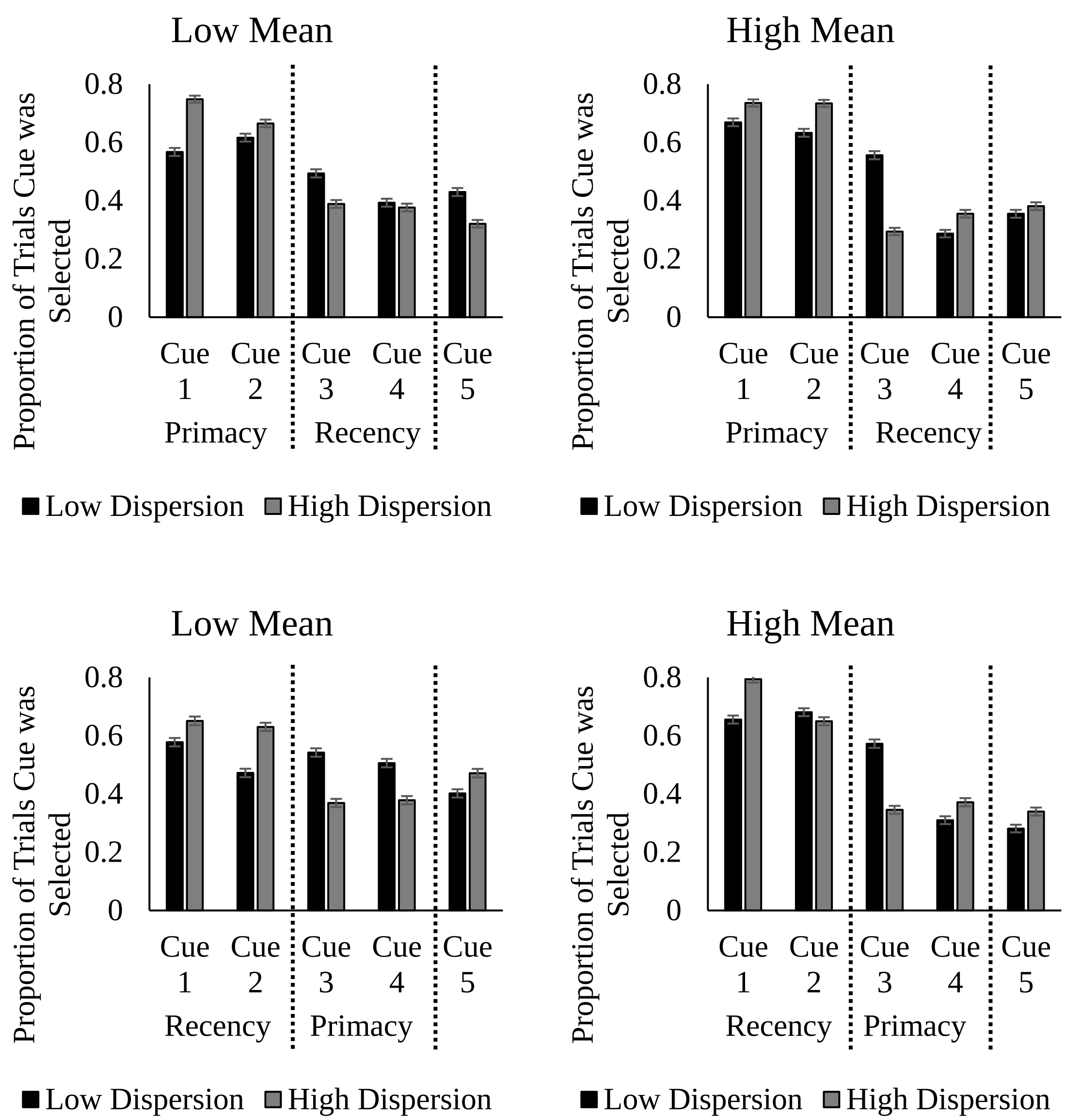

Figure 5. Proportion of trials each cue was selected by dispersion and mean conditions. Note that Cue 1 and Cue 2 had the same validity by Cue 2 correctly discriminated more at the beginning (primacy) or end (recency) and that Cue 3 and Cue 4 had the same validity, but Cue 4 correctly discriminated more at the end (recency) or beginning (primacy). Bars indicate standard error.

Tests of the preference for the primacy cue, did not find a significant effect of order condition (χ^2 (1) = 1.22, p = .269), dispersion of validity (χ^2 (1) = 1.37, p = .242), mean validity (χ^2 (1) = 1.03, p = .311) or the number of cues participants were instructed to select (χ^2 (1) = 3.24, p = .357). There was a significant interaction between order condition and dispersion of validity (χ^2 (1) = 4.56, p = .03). Post-hoc analyses indicated that those in the condition with primacy manipulated for the less valid cue pair and low dispersion of validity were much less likely to select the primacy cue (O = -.61) compared to those in the same order condition but with high dispersion of validity (z = 2.35, p = .019; O = .05) and compared to those with low dispersion but primacy manipulated for the most valid cue pair (z = 2.33, p = .02; O = .05). No other conditions differed significantly. There was also a significant interaction between mean validity and dispersion of validity (χ^2 (1) = 4.63, p = .03), such that those in the high mean condition with low dispersion of validity were much less likely to select the primacy cue (O = -.59) compared to those in the high mean condition with high dispersion of validity (z = 2.35, p = .019; O = .06) and compared to those in the low mean condition with low dispersion of validity (z = 2.23, p = .026; O = .03). No other conditions differed significantly.

A chi-square goodness-of-fit was conducted to compare the selection of only the primacy cue to its similar validity middle cue when only one of them was selected, collapsed across the number of cues selected. This was done to further examine the preference between the primacy cue and the middle cue within participants. Selection of either member of the pair was found to significantly differ from a chance selection of .5 with preference for the middle cue when primacy was manipulated for the best cue pair and dispersion was high and mean was low, when primacy was manipulated for the most valid cue pair and dispersion was low and mean was high, and when primacy was manipulated for the less valid cue pair and dispersion was low and mean was high, as shown in Table 7. Preference for the primacy cue was found in the condition in which primacy was manipulated for the best cue pair and the dispersion was low and mean was low.

Table 7
Chi-square Goodness-of-fit Tests for Preference for Primacy Cue

Order	Dispersion	Mean	Preference for Primacy Cue	Chi-Square	*P*
Primacy/Recency	High	High	0.4981	.00076 (1)	0.93
Recency/Primacy	High	High	0.5309	1.91 (1)	0.17
Primacy/Recency	High	Low	0.4072	20.24 (1)	< .0001
Recency/Primacy	High	Low	0.5097	.21 (1)	0.64
Primacy/Recency	Low	High	0.4573	3.84 (1)	0.05
Recency/Primacy	Low	High	0.2597	149.49 (1)	<.0001
Primacy/Recency	**Low**	**Low**	**0.5542**	**6.83 (1)**	**0.009**
Recency/Primacy	Low	Low	0.46	3.44 (1)	0.06

Tests of the preference for the recency cue also did not find a significant effect of order condition (χ^2 (1) = 0.17, p = .678), dispersion of validity (χ^2 (1) = 2.69, p = .10), mean validity (χ^2 (1) = .08, p = .775) or the number of cues participants were instructed to select (χ^2 (1) = 2.81, p = .423). There was a significant interaction between of the dispersion of validity and order condition (χ^2 (1) = 6.86, p = .009) and a significant three-way interaction between mean validity, dispersion, and order condition (χ^2 (1) = 8.05, p = .005). Post hoc analyses showed that when recency was manipulated for the less valid cue pair, mean validity was high, and dispersion was low, participants were much less likely to select the recency cue (O = -1.09) compared to when recency was manipulated for less valid cue pair, mean validity was high, and dispersion was high (z = 3.26, p = .001; O = .26) and compared to when recency was manipulated for more valid cue pair, mean validity was high, and dispersion was low (z = 3.20, p = .001; O = .16).

A chi-square goodness-of-fit was conducted to compare the selection of the recency cue to its similar validity middle cue when only one of them was selected, collapsed across the number of cues selected as shown in Table 8. Selection of either member of the pair was found to significantly different from a chance selection of .5 such that participants tended to prefer the middle cue when recency was manipulated for the most valid cue pair and both mean validity and dispersion were high and when both mean validity and dispersion were low. Participants also tended to prefer the middle cue when recency was manipulated for the less valid cue pair and dispersion as low for both levels of mean validity. Selection of either member of the pair was found to significantly different from a chance selection of .5 such that participants tended to prefer the recency cue over the middle cue when dispersion was high, mean was high, and recency was manipulated for the less valid cue pair.

Table 8

Chi-square Goodness-of-fit Tests for Preference for Recency Cue

Order	Dispersion	Mean	Preference for Recency Cue	Chi-Square	*P*
Primacy/Recency	**High**	**High**	**.5766**	**12.26**	**.0005**
Recency/Primacy	High	High	.3353	56.34	<.0001
Primacy/Recency	High	Low	.4868	.42	.51
Recency/Primacy	High	Low	.4765	1.13	.29
Primacy/Recency	Low	High	.2565	163.62	<.0001
Recency/Primacy	Low	High	.5314	1.88	.17
Primacy/Recency	Low	Low	.4009	25.56	<.0001
Recency/Primacy	Low	Low	.3827	29.32	<.0001

5.1.2.3 Decision Outcome

A generalized linear model with a binomial logit link was also used to test the effect of order condition, dispersion of validity, and mean validity on participant's decisions. This analysis was conducted on trials in which participants had access to all cues and compensatory and noncompensatory strategies differed in which company should be selected. Decisions differed significantly between dispersion conditions (χ^2 (1) = 9.34, p = .002), mean condition (χ^2 (1) = 5.99, p = .01), and order condition (χ^2 (1) = 7.37, p = .007). There was also a significant three-way interaction between order condition, mean condition, and dispersion condition (χ^2 (1) = 4.74, p = .029). In general, participants in the low dispersion group (*O* = 1.69) were more likely to select the compensatory option compared to those in the high dispersion group (*O* = 1.12). The effects of the order condition and the mean condition depended on the three-way interaction. Participants in the high mean group were more likely to select the compensatory option compared to those in the low mean group, except when primacy was manipulated for the most valid cue pair and the dispersion was high as shown in Figure 6. For the effect of serial position, those with primacy manipulated for the least valid cue pair tended to select the compensatory option more frequently than when primacy was manipulated for the most valid cue pair except for in the low dispersion and high mean condition in which the difference disappears.

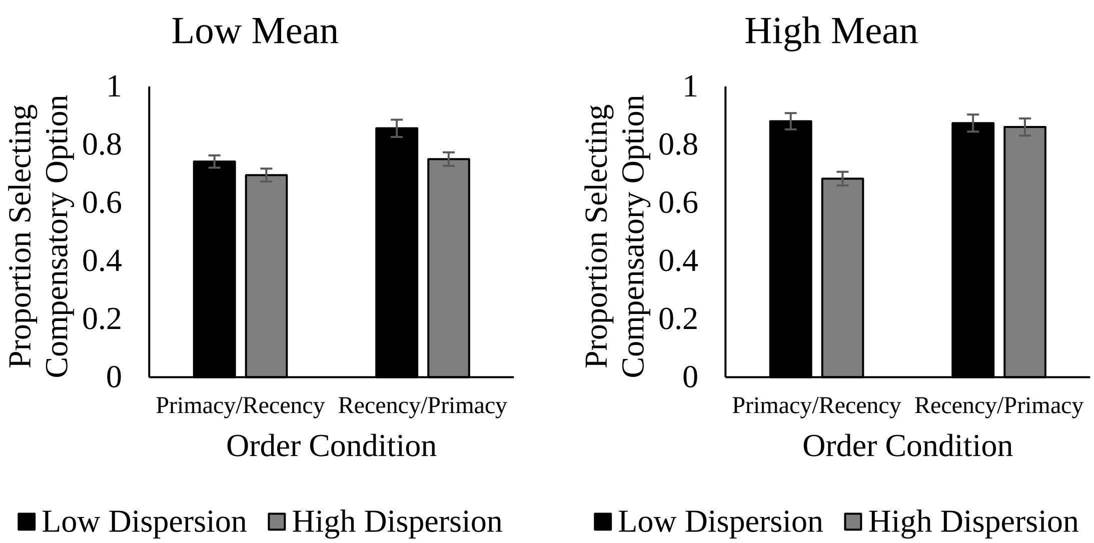

Figure 6. Proportion selecting the compensatory option by order condition, mean condition, and dispersion condition. Bars indicate standard errors.

5.1.3 Discussion

In general, this study provided partial support for the hypotheses of an accessibility-based framework for cue-based inferences. The goal of this study was to test predictions based on the accessibility framework for the effects of manipulations of validity and serial position. In terms of decision outcome, the manipulations of serial position, dispersion of validity, and mean validity affected the likelihood of selecting the compensatory option. As predicted, participants in the low dispersion group were more likely to select the compensatory option compared to those in the high dispersion group. This replicates prior research that has found that people are sensitive to the dispersion of cue validities when making decisions (Bröder, 2003; Newell & Shanks, 2003). Although there were differences in training performance between these conditions, these differences were small, meaning learning differences likely do not account for the above effect.

Moreover, those with primacy manipulated for the less valid cue pair also tended to select the compensatory option more frequently, except in one condition. Again, this is not likely caused by the small differences in performance at training. This partially supports the hypothesis that there would be more compensatory-type of behavior when primacy was manipulated in the less valid cue pair. Additionally, the high mean and high dispersion condition with primacy manipulated for the best cue pair showed the least amount of compensatory decisions, as predicted. It also provides further evidence that serial position manipulations interacted with other aspects of the decision environment. The manipulation of the serial order did not seem to affect decisions in the environment with low dispersion low and high mean. This could be because there was already a strong preference for the compensatory option in the low dispersion and high mean condition in general. Overall the results suggest that manipulations of serial position can affect how people combine cue information to arrive at a decision.

Although mean validity was not predicted to have a main effect on decisions, the results also indicated that those in the high mean condition were more likely to select the compensatory option compared to those in the low mean condition, except in one condition. It could be the case that more valid cues were given relatively more weight when mean validity was low because the less valid cues had very low validity, resulting in fewer compensatory decisions. For example, in the low mean and high dispersion condition, the least valid cue only performed slightly above chance. This effect was absent when primacy was manipulated for the most valid cue pair and the dispersion was high. It is possible that the manipulation of serial order increased noncompensatory decisions in this condition, making the high and low mean conditions more similar. This is partially consistent with the hypothesized effect that there would be fewer compensatory decisions in the high mean and high dispersion condition overall.

It was hypothesized that the effects of serial position would interact with manipulations of validity, such that the primacy effect would be stronger under certain conditions. Although this study found evidence for an interaction between cue validity and serial position effects, the pattern of the interaction in this experiment was difficult to interpret and not consistent with predictions from the accessibility-based framework. It was hypothesized that serial position effects would likely be stronger in conditions in which the dispersion of validity was low. Yet, preferences for the primacy cue were not consistently found when the dispersion of validity was low. In fact, in three out of the four conditions with low dispersion of validity, there was a trend for preferring the cue that discriminated correctly more often in the middle for both primacy and recency cue pairs.

In general, the expected differences in cue preferences were not well-supported by the results. There was not a consistent pattern for either a primacy effect or a recency effect. A primacy effect was only found in the condition with low mean validity and low dispersion and only when primacy was manipulated for the most valid cue pair. Similarly, a recency effect was only found in one condition: recency manipulated for the less valid cue pair with high dispersion and high mean validity. This is not consistent with previous research which found a robust primacy effect when primacy was manipulated for the less valid cue pair and dispersion of validity was low and found no evidence for a recency effect (Lawrence et. al., 2018b). However, the mean validity and dispersion of validity in the previous experiment fell between the manipulated values in the current study. It is possible that the effect interacts with validity in a nonlinear manner.

Moreover, in a few conditions, there was evidence for a preference for the middle cue over the primacy cue. For the recency manipulation, this pattern was even stronger. In fact, in half of the conditions, there was evidence for a preference for the middle cue over the recency cue. This is not completely inconsistent with previous research (Lawrence et al., 2018b), which found some evidence for a preference for the middle cue over the recency cue. However, in that paper, the validity of the middle cue was slightly higher than the validity of the recency cue so, unlike in the current experiment, validity could not be ruled out as a possible explanation.

In order to better understand the effects of serial position, future research could focus on measuring additional process level data. For example, eye-tracking data might help elucidate whether there are differences in attention within learning blocks. It was hypothesized that cues that correctly discriminated early in blocks would have an encoding advantage. However, changes in attention could explain the preference of the middle cue in some conditions. It is possible that participants were more attentive during certain parts of the learning blocks, resulting in preferences for one cue over another even when the cues had the same validity. These attentional differences could have interacted with the manipulations of mean validity and dispersion of validity such that they only operated under certain conditions.

Importantly, none of the current accounts of cue-based inferences can explain why participants would prefer the middle cue in some situations. In this experiment, cues were paired such that they had the same validity and discrimination rates but differed in terms of the serial position of correct discriminations. Thus, validity-based accounts (e.g. adaptive toolbox, EAM, and PCS) would predict that participants should show no preferences between the members of each cue pair. Alternatively, a memory-based account would predict that participants should show a preference for the cues that correctly discriminated more at the beginning or end of each block, but not in the middle. Even though the results of this experiment challenge all frameworks for cue-based inferences discussed in this dissertation, the effects of the manipulations on decision outcomes provide partial support for the accessibility-based framework.

5.2 Experiment 2: Effect of Delay on Cue-Based Inferences

Although the first experiment tested the accessibility framework's predictions about the effect of manipulations of cue validity and memory on cue preferences and decision, it did not test predictions about search order and stopping behavior. This second experiment tests the accessibility framework by examining search behavior more closely. Specifically, in this experiment participants were able to search cues sequentially and decide for themselves when to stop search. This provides a further test of the accessibility model and allows for model comparisons between the HyGene based accessibility framework and the other frameworks discussed in this dissertation, except PCS which does not have a specified search process.

Memory was manipulated differently in this study compared to Experiment 1. Rather than manipulating the serial position of correct discriminations, this study introduced a retention interval between the training phase and the final test phase. Effects of retention intervals, the time between learning and recall, on memory are well established. Although the exact mechanisms leading to forgetting are under debate, there is clear evidence that the more time that passes between learning and test, the less information is retrieved (for review see Wixted, 2004). However, the current frameworks for cue-based inferences cannot account for such a pervasive memory phenomenon.

Because the accessibility framework is a memory-based model, only a slight adjustment is needed to allow it to account for the effects of retention intervals. Within MINERVA-2 (Hintzman, 1988), the model upon which HyGene is based, decay could only be modeled using the learning parameter (*L*). However, this implementation for decay has been criticized for being unconstrained (Dimov, 2017). Instead of using L to model decay, the accessibility model can implement decay as a power function such that activation after a delay is a function of initial cue activation and a decay parameter:

$$A_{sd} = A_s t^{-(A_s * d)} \tag{1}$$

where A_{sd} is the activation of a cue after a delay, A_s is the initial cue activation, t is the time between learning and retrieval, and d is the decay parameter. Although other functions accounting for decay exist (Rubin & Wenzel, 1996), the power function was chosen because it has been shown to successful (Wixted & Carpenter, 2007; Wixted & Ebbesen, 1991) and because it has been used in other cognitive architectures, such as ACT-R (Anderson et al., 2004). Modeling decay rate as a function of initial activation is based on the way practice effects have been implemented in ACT-R (Pavlik & Anderson, 2005). In contrast to the accessibility framework, effects of a delay cannot be easily modeled in the other frameworks because they assume cue search is based on pre-computed hierarchies, which are not memory-based.

As noted above, the overall goal for this dissertation is to show that the accessibility framework accounts for differences in decisions processes resulting from both direct memory manipulations and manipulations of the decision environment. Thus, in addition to manipulating memory via different retention intervals, the dispersion of cue validity was also manipulated. This manipulation was similar to Experiment 1 with one condition having highly dispersed cue validities and one condition having less dispersed cue validities, which was manipulated between participants. The retention interval was manipulated within participants with a test phase immediately following learning and another test phase a week later. These conditions were compared in terms of cue search (both cue selection and order of selection), stopping behavior (number of cues searched, decision time), and decision (option selected).

The accessibility framework makes a number of predictions in regards to the effect of these manipulations on cue-based inferences. As shown in the previous experiment, it is expected that participants in the less dispersed validity condition will show more compensatory like behavior compared to the highly dispersed condition. Participants should search more cues, show a weaker preference for the most valid cue (as measured by selection and order of selection), and select the compensatory option more often. Although the mechanisms differ, the other models make similar predictions in regards to the effects of cue dispersion. The evidence accumulation model assumes the decision threshold will be reached with a less extensive search in the high dispersion condition compared to the low. This means the same pattern of behavior predicted for the accessibility framework is also predicted by EAM. The adaptive toolbox also assumes differences in behavior between the groups, such that participants may use Take-the-Best in the high dispersion condition and WADD in the low dispersion condition. This means in the low dispersion condition the adaptive toolbox assumes exhaustive search, which neither the accessibility framework nor EAM assume. Note, parallel constraint satisfaction does not specify search processes but assumes all available information is used.

Although all models assume differences caused by the manipulation of the dispersion of cue validity, only the accessibility framework can account for differences caused by the retention interval. The accessibility framework assumes participants will behave differently during the test phase after the week-long retention interval compared to the immediate test phase. This may interact with the manipulation of the dispersion of cue validity. The retention interval should result in more compensatory-type of behavior in both dispersion groups, but the effect may be stronger in the high dispersion group. Before getting into specifics of the experiment, simulations of the different decision models will be discussed to demonstrate these predictions.

5.2.1 Model Simulations

5.2.1.1 Accessibility-Based Framework

Simulations were conducted to demonstrate the behavior of the HyGene accessibility-based model with the decay parameter under two different validity dispersions. The two data sets used for training participants in the current experiment were used as training data for the HyGene model. The model was trained using the method described in the simulation section of this dissertation. Additionally, the decay parameter detailed in the introductory section for this experiment was used in this simulation. In addition to training on either the high or low validity dispersion data sets, the encoding parameter and the decay parameter were also manipulated within the model. The encoding parameter was manipulated to show the differences in model performance under different levels of learning. The decay parameter was manipulated to show the effects of different decay rates on the performance of the model, as would be expected after a retention interval. The discussion of the results of the simulation will focus on the effects of decay and how it interacts with the encoding parameter and the dispersion of validity. A detailed discussion of the effects of the encoding parameter and the dispersion of validity on the model can be found in the previous simulation.

5.2.1.1.1 Number of Cues Generated

In general, the number of cues generated increased as the dispersion of validity decreased (as shown in Table 9). This pattern was stronger for higher levels of the encoding parameter. As the encoding parameter decreased, the number of cues selected tended to decrease. Yet, the number of cues generated increased as the decay parameter increased, matching the prediction that more cues are used when there is a delay between training and test.

Table 9
Average Number of Cues Selected by Data Set, Encoding Parameter, and Decay Parameter

	High Dispersion			Low Dispersion		
	Decay = 0	Decay = 1	Decay = 2	Decay = 0	Decay = 1	Decay = 2
Encoding						
1.0	1.74	1.90	2.00	1.96	2.04	2.15
0.6	1.76	1.91	2.01	1.93	2.04	2.12
0.2	1.72	1.89	1.98	1.75	1.92	2.03

In addition to looking at the average number of cues generated, the performance of the model under different parameters was also compared in terms of the proportion of decisions based on one cue, see Table 10. As the decay parameter increased, the proportion of decisions based on a single cue also decreased. Again, demonstrating the prediction that people use more cues when there is a delay between training and test.

Table 10
Proportion of Decisions Based on a Single Cue by Data Set, Encoding Parameter, and Decay Parameter

	High Dispersion			Low Dispersion		
	Decay = 0	Decay = 1	Decay = 2	Decay = 0	Decay = 1	Decay = 2
Encoding						
1.0	.41	.33	.29	.32	.28	.24
0.6	.41	.33	.29	.33	.28	.25
0.2	.43	.33	.30	.42	.32	.28

5.2.1.1.2 Cue Preferences

When the decay parameter was set to zero, the results of this simulation matched the simulation discussed earlier in this dissertation as shown in Figure 7. In general, as the decay parameter increased, preference for the most valid cue decreased and preference for the least valid cue increased. At the lowest level of the encoding parameter, preferences between cues were nearly equal, especially when dispersion was low. At higher levels of the encoding parameter, preference for the middle cues tended to increase, especially for the high dispersion condition. Overall, these simulations demonstrate the prediction that cue preferences change as a result of the delay between learning and test.

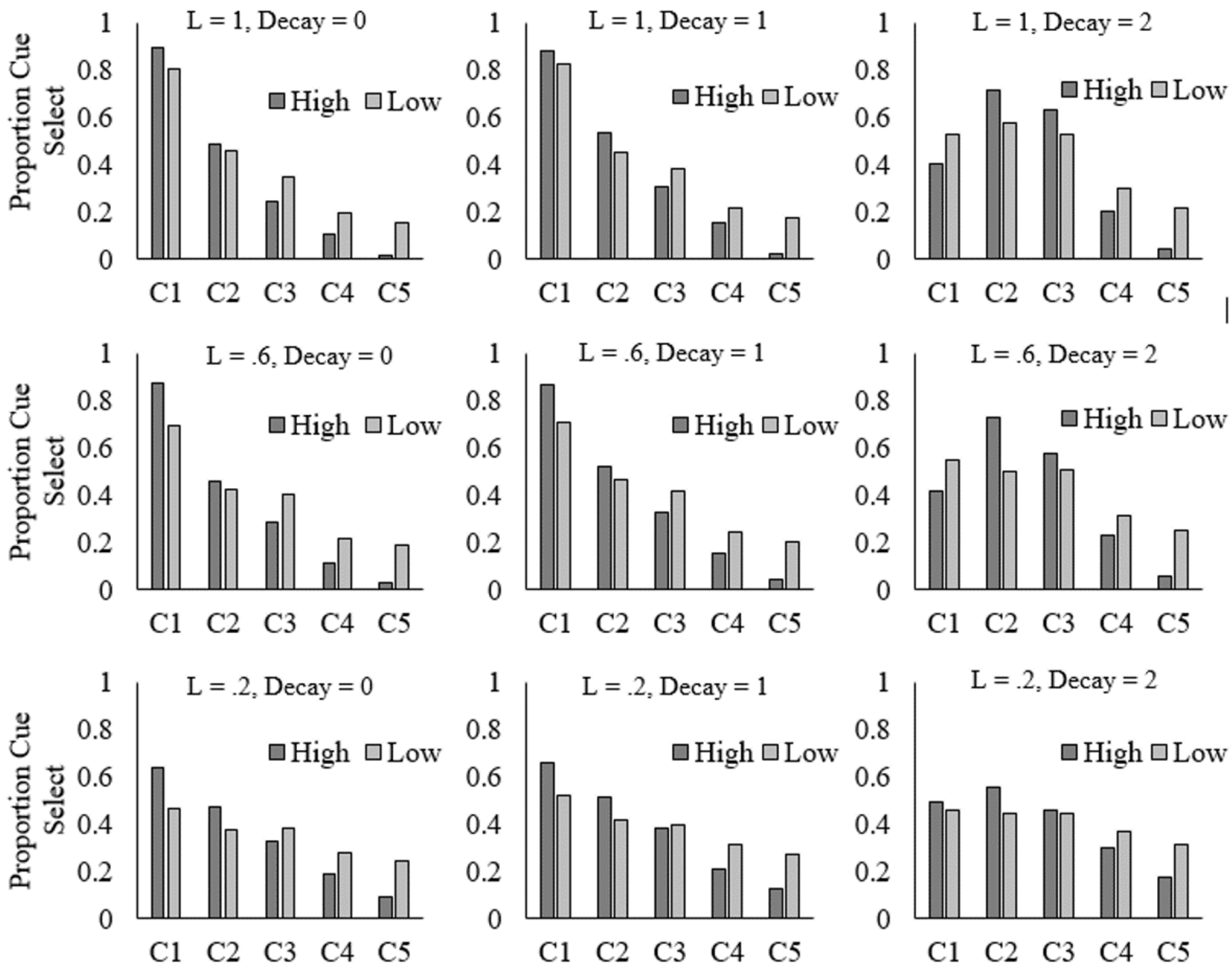

Figure 7. Proportion of trials each cue is selected by validity dispersion, learning parameter, and decay parameter.

5.2.1.1.3 Decision Outcome

When the decay parameter was high for moderate and high levels of the encoding parameter, the proportion of decisions matching compensatory strategy was increased (shown in Table 11). This demonstrates the prediction that the number of compensatory decisions increases when there is a delay between learning and test. Decay did not have a strong effect on the option selected when the decay parameter was moderate or when the encoding parameter was low.

Table 11
Proportion of Decision Matching Compensatory Rule When Compensatory and Noncompensatory Differ by Data Set, Encoding Parameter, and Decay Parameter

	High Dispersion			Low Dispersion		
	Decay = 0	Decay = 1	Decay = 2	Decay = 0	Decay = 1	Decay = 2
Encoding						
1	.250	.251	.456	.349	.342	.508
.6	.270	.275	.474	.437	.431	.528
.2	.438	.439	.438	.567	.562	.620

5.2.1.2 Simulation of EAM, TTB, and WADD

Simulations were also conducted to show the behavior of the other models: TTB, WADD, and EAM. Because none of these models include a learning portion, the models were based on the ecological validity of the two dispersion conditions. Model behavior was simulated for the test trials. This is straightforward for both TTB and WADD. For TTB, cues were searched in order of validity until a discriminating cue was found for each test trial. The model then selected the option with the higher value on that cue. For WADD, every cue was searched for each trial and the option selected was determined by weighing the cues by their validity.

The behavior of EAM was simulated following the description of EAM found in Cummins & Newell (2005). The likelihood of each option being higher on the criterion was calculated based on the validity of the cues. This was done sequentially until the likelihood exceeded the threshold parameter. For example, at a threshold of 0, the model mimics TTB such that the option that is highest on the first discriminating cue is selected. Simulations were run for five different levels of the threshold parameter: 0, 1, 2, 3, and 4. Because EAM, TTB, and WADD, are deterministic models, they were only simulated on a single run through the test trials.

5.2.1.2.1 Number of Cues Selected.

For EAM, as the threshold increased the average number of cues selected increased as shown in Table 12. This increased more quickly for the low dispersion condition than for the high dispersion condition. For TTB, the average number of cues selected was the same as EAM with a threshold of 0, 1.64. WADD always selected all cues. Compared to HyGene accessibility framework, EAM generally selected a larger number of cues.

Table 12
Average Number of Cues Selected for EAM at Different Levels of the Threshold Parameter

Threshold	0	1	2	3	4
High	1.64	1.94	2.56	4.38	4.9
Low	1.64	2.54	4.38	4.92	5

As shown in Table 13, at low levels of the threshold parameter the majority of decisions were based on single cue, but this dropped to zero at higher levels of the threshold parameter for EAM. At low levels of the threshold parameter, EAM predicts a larger number of decisions based on a single cue than the HyGene accessibility-based framework. HyGene predicts more cues based on a single cue compared to EAM at higher levels of the threshold parameter.

Table 13
Proportion of Decisions Based on a Single Cue for EAM at Different Levels of the Threshold Parameter

Threshold	0	1	2	3	4
High	.6	.6	.6	0	0
Low	.6	.6	0	0	0

5.2.1.2.2 Cue Preferences

Unlike the HyGene accessibility framework, the other models showed less variability in cue selection behavior; these models always selected the most valid cue as shown in Figure 8. As the threshold increased for EAM, the proportion of trials using the less valid cues also increased. Again, TTB is the same as EAM at a threshold of 0, meaning it used the best cue almost exclusively. In contrast, WADD always used all cues.

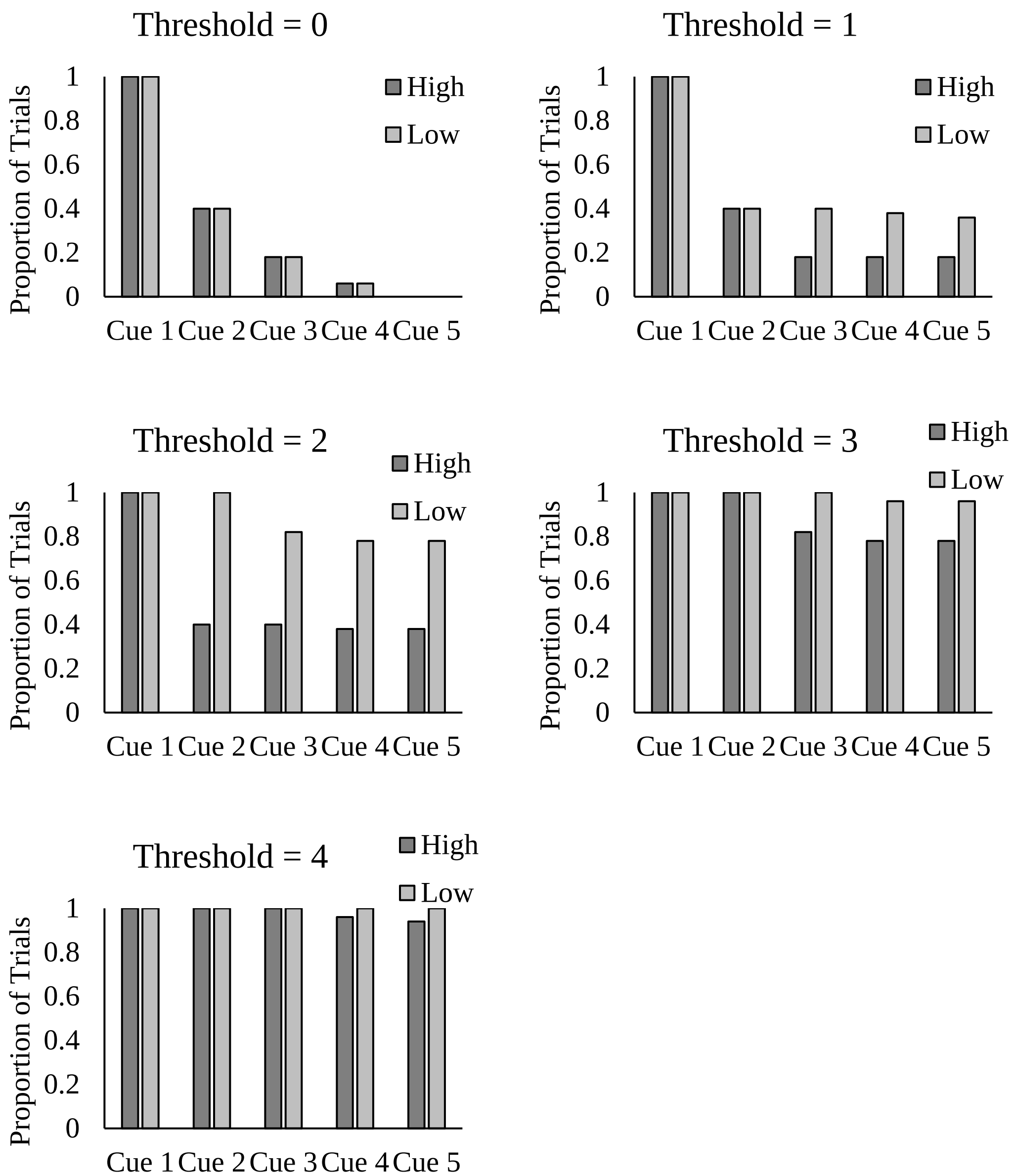

Figure 8. Proportion of trials each cue is selected by threshold parameter and dispersion of validity.

5.2.1.2.3 Decision Outcome

Compared to HyGene, the other models had less variability in the option selected (see Table 14). No matter the threshold, EAM for the high dispersion condition always selected the non-compensatory option because high values on the less valid cues could not out-weigh a high value on the more valid cues. For the low dispersion condition, EAM always selected the compensatory option except at low levels of the threshold parameter. TTB always selected the non-compensatory option for both dispersion conditions and WADD always selected the compensatory option.

Table 14.
Proportion of Decisions Matching Compensatory Rule When Compensatory and Noncompensatory Differ by Dispersion of Validity and Threshold Parameter

Threshold	0	1	2	3	4
High	0	0	0	0	0
Low	0	0.4167	1	1	1

5.2.2 Method

There were 77 total participants in this study. However, only those who attended both sessions were included in data analyses. This resulted in 66 participants with 34 in the high dispersion condition and 32 in the low dispersion condition. Participants received course credit for their participation. They also received a bonus check based on the points they earned throughout both sessions of the experiment. The average bonus earned in the first session was $4.19 and in the second session it was $2.53

The design of this experiment was a 2 (retention interval) by 2 (dispersion of cue validity) with the retention interval manipulated within subjects and dispersion manipulated between subjects. The two conditions for the dispersion of cue validity were a standard deviation of .05 (low) or .14 (high).

The training phase of this experiment was like Experiment 1 with participants learning how five different cues could be used to predict stock price. The procedure differed from Experiment 1 in a few ways. The dispersion of the cue validities was manipulated so that in one group the standard deviation of the cue validities was .05 and in the other group the standard deviation of the cue validities was .14 as shown in Table 15. The training phase was not divided into blocks, instead, participants completed 100 training trials in a single block. To incentivize good performance, participants earned 5 points for each correct answer during this phase.

Table 15
Validity of the Cues by Condition for Experiment 2

	High Dispersion	Low Dispersion
Cue 1	.91	.77
Cue 2	.78	.73
Cue 3	.70	.70
Cue 4	.62	.67
Cue 5	.51	.63

Table 15 (continued).

Mean	.70	.70
SD	.14	.05

Once participants completed the training phase, they began the first test phase. Unlike Experiment 1, participants were allowed to determine for themselves the number of pieces of information to use and when they would like to stop search. The test phase looked similar to the training phase, except none of the cue information was immediately available. For each cue, participants were able to reveal information by clicking on a button to purchase the cue information for both companies. Participants were required to purchase at least one piece of information. To incentivize more careful cue search, each cue cost 1 point to be revealed. However, participants earned 10 points for each correct answer, so they could still earn 5 points for a correct response even if they decided to purchase all the information. Participant s completed 50 test trials during this phase. A second test phase then took place a week later. This phase was the same as the previous test phase.

5.2.3 Results

5.2.3.1 Learning

A generalized linear model with a binomial logit link was used to test if trial and dispersion of validity affected performance during the training trials. Trial (χ^2 (1) = 5.90, p = .015) significantly affected performance during training such that performance increased across training trials. Dispersion of validity did not significantly affect performance during training (χ^2 (1) = 2.20, p = .138) and there was no interaction between dispersion of validity and trial (χ^2 (1) = 0.24, p = .624). A binomial test, collapsed across conditions, also showed that the proportion of correct responses on the last quarter of training trials (.82) was above a chance level of .5 (z = 25.90, p <.0001), as shown in Figure 9, suggesting that participants were able to learn.

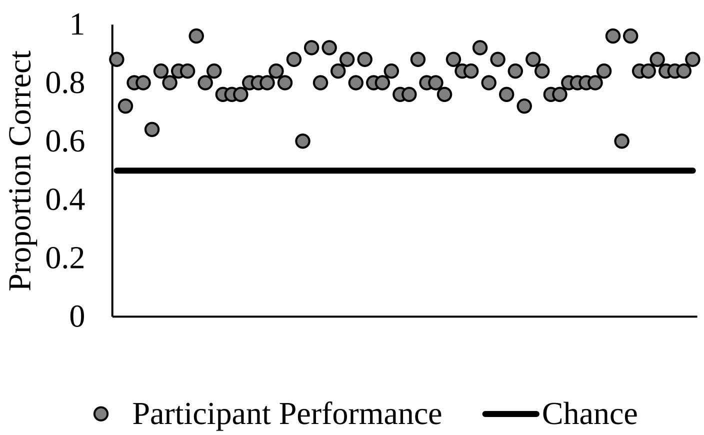

Figure 9. Proportion correct on the last 25 trials by participant.

5.2.3.2 Number of Cues Selected and Decision Time

A Poisson regression with a log link was used to test whether dispersion of validity and retention interval affected the number of cues participants selected at test. Neither dispersion of validity (χ^2 (1) = 2.85, p = .091) nor retention interval (χ^2 (1) = 0.58, p = .445) significantly affected the number of cues selected and there was no interaction (χ^2 (1) = 0.04, p = .851). The median number of cues selected per trial was 4 as shown in Figure 10.

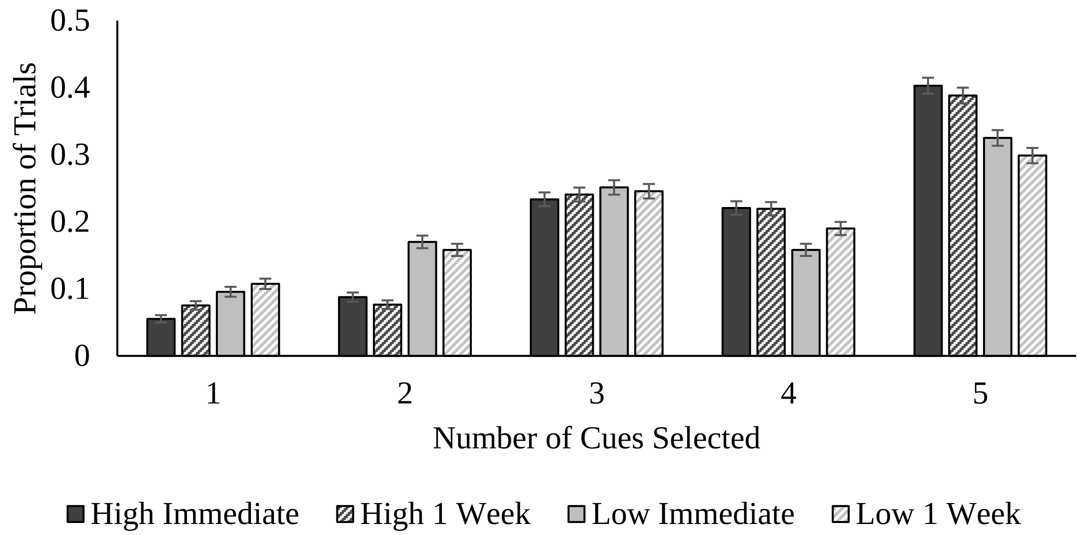

Figure 10. Number of cues selected by dispersion condition and retention interval. Bars indicate standard errors.

A generalized linear regression was used to test whether dispersion of validity and retention interval affected the amount of time participants took for each trial at test. Decision time was log transformed. Dispersion of validity did not significantly affect decision time (χ^2 (1) = 2.06, p = .15). Retention interval significantly affected decision time (χ^2 (1) = 26.10, p < .001) such that participants took less time in the delayed session (M = 6.013, SE = .335) compared to the session immediately following training (M = 7.51, SE = .35).

5.2.3.3 Decision Outcome

A generalized linear model with a binomial logit link was used to test the effect of dispersion of validity and retention interval on participant's decisions. This analysis was conducted on trials in which the compensatory and noncompensatory strategies differed in which company should be selected. The number of cues selected was also included as a predictor in the model. Retention interval did not significantly affect the option selected (χ^2 (1) = .08, p = .777). Dispersion condition significantly affected the likelihood of selecting the compensatory option (χ^2 (1) = 6.46, p = .011) such that participants were more likely to select the compensatory option in the low dispersion condition (O = 2.209; 82.94%) compared to the high (O = 1.540; 77.7%). Moreover, participants were also more likely to select the compensatory option as the number of cues selected increased (χ^2 (1) = 20.47, p < .001) as shown in Figure 11.

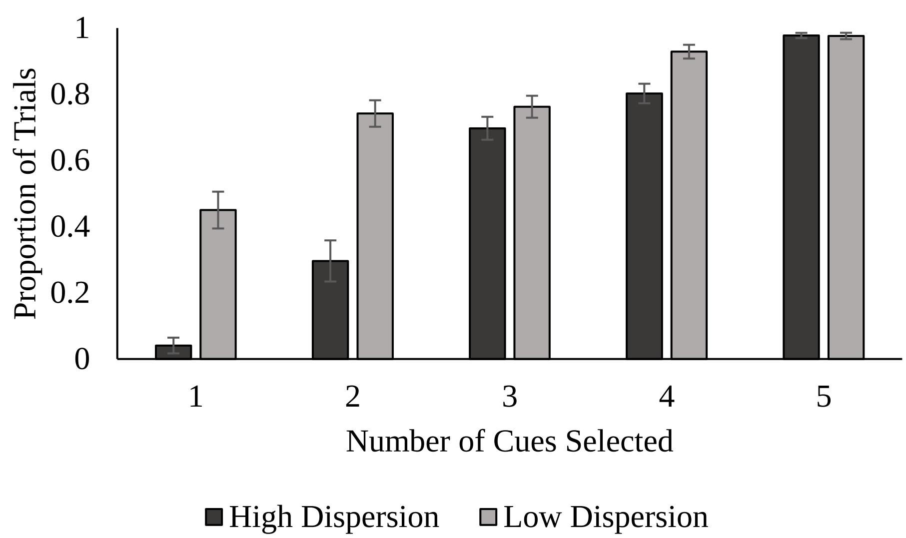

Figure 11. Proportion selecting the compensatory option when compensatory and noncompensatory differ. Bars indicate standard errors.

A generalized linear model with a binomial logit link was also used to test the effect of dispersion of validity and retention interval on participant's performance during the test trials. This analysis was conducted on trials in which the compensatory and noncompensatory strategies indicated the same company should be selected. The number of cues selected was also included as a predictor in the model. Neither dispersion of validity (χ^2 (1) = 1.45, p = .229) nor number of cues selected (χ^2 (1) = 1.77, p = .183) significantly affected the option selected. Retention interval did significantly affect the likelihood of selecting the correct option (χ^2 (1) = 5.88, p = .015) such that participants were more likely to select the correct option in during the immediate session (O = 1.318; 78.83%) compared to the delayed session (O = 1.164; 76.20%).

5.2.3.4 Cue Preferences

A generalized linear model with a binomial logit was also used to test whether cue preferences were affected by dispersion of validity and retention interval (see Figure 12 for the overall pattern of selections). This was done individually for each cue with the number of cues selected also included as a predictor. The number of cues selected was the only significant predictor for the best cue (χ^2 (1) = 18.98, p < .001), the second best cue (χ^2 (1) = 28.51, p < .001), and the worst cue (χ^2 (1) = 44.58, p < .001), such that the more cues selected the more likely a participant was to select that particular cue.

For the third best cue, there was a significant interaction between number of cues selected and dispersion condition (χ^2 (1) = 4.78, p = .029) such that when few cues were selected (1 or 2), those in the low dispersion condition were more likely to select this cue, but when more cues were selected (3 or 4), those in the high dispersion condition were more likely to select this cue. Number of cues selected was also significant for the third cue (χ^2 (1) = 38.88, p < .001). Dispersion condition affected the selection of the fourth most valid cue (χ^2 (1) = 4.64, p = .031) such that those in the low dispersion condition were more likely to select this cue (O = .988) compared to those in the high dispersion condition (O = 1.916). Number of cues selected was also significant for the fourth cue (χ^2 (1) = 33.07, p < .001).

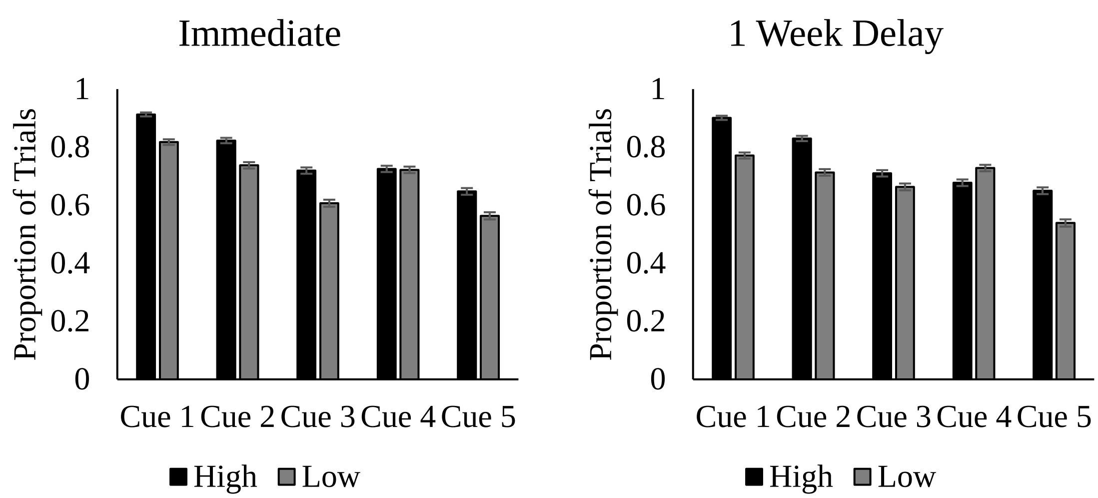

Figure 12. Proportion of trials each cue was selected by dispersion condition and retention interval. Bars indicate standard errors.

Figure 13 shows the percentage of trials in which each individual cue was selected in a particular order position out of all trials in which it was selected, collapsed across conditions. A generalized linear model with a cumulative logit link was used to test whether the order of cues selected was affected by dispersion of validity and retention interval. This was done individually for each cue with the number of cues selected also included as a predictor. These analyses were only run on trials in which the cue being modeled was selected. Number of cues significantly predicted the order of cue selections for all cues (χ^2 (1) = 16.69, p < .001; χ^2 (1) = 18.11, p < .001; χ^2 (1) = 24.07, p < .001; χ^2 (1) = 19.80, p < .001, χ^2 (1) = 21.31, p < .001). This was an artifact of the nature of the order of selection because the more cues that were selected later a cue could be selected in the order. The number of cues selected was included in the model to control for this when looking at the effects of the other manipulations.

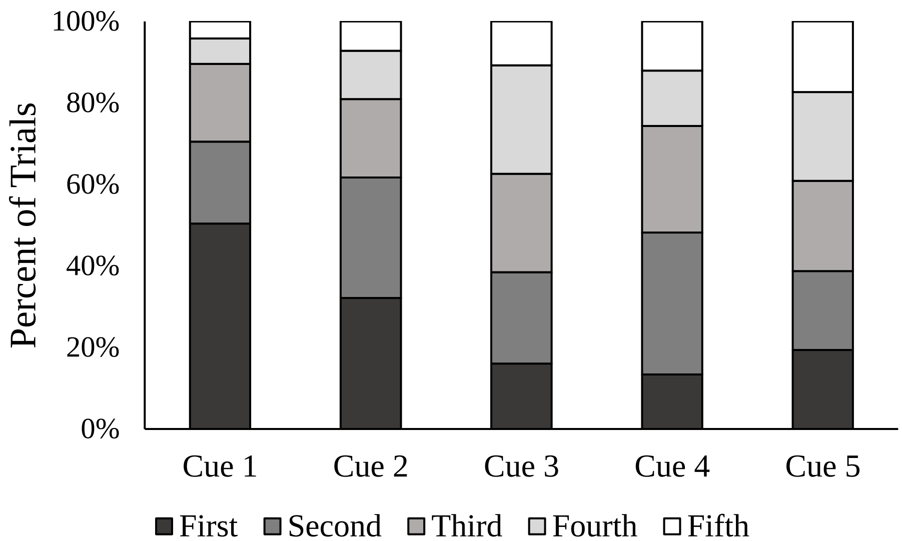

Figure 13. Percent of trials for each order of selection by cues in order of validity. For example, Cue 1 was selected first in approximately 50% of the trials in which it was selected.

For the third cue, retention interval significantly affected the order of cue selections (χ^2 (1) = 4.88, p = .027) such that the third cue tended to be selected earlier in the second session compared to the first. For the fourth cue, there was a significant interaction between the number of cues selected and retention interval (χ^2 (1) = 4.03, p = .045). When fewer than five cues were selected, there were no large differences in the selection order. Yet when all five cues were selected, the fourth cue tended to be selected more in middle during the immediate test phase but more often as either the first of the last cue in the delayed session. No other manipulations were significant for the other cues.

5.2.3.5 Model Fitting

In order to compare how well each model accounted for the data, fit analyses were conducted by determining the likelihood of participant cue selection behavior under the various models: random, TTB, WADD, EAM, and HyGene accessibility-based framework. The probability of each cue being selected was determined by normalizing the cue selections of the various models based on the simulations of each model discussed above. The response probabilities were applied stochastically without replacement to calculate the likelihood of the participant's pattern of cue selections. For example, assume a participant selected Cue 3 and then Cue 4. Under a random model, the likelihood of selecting Cue 3 first is .2 and the likelihood of selecting Cue 2 second is .25; this results in a likelihood of .05 (.2 multiplied by .25). The probability of each number of cues being selected was also determined by normalizing the proportion of decisions based on each number of cues from the simulated model behaviors. For example, the probability of choosing any number of cues in a random model is .2. The log likelihood of the pattern of cues selected was combined with the log likelihood of the number of cues selected to calculate the log likelihood of each model for each trial for individual participants. These likelihoods were then used to derive fit statistics: G^2 and BIC. This was done at the individual participant level and at the aggregate level for both sessions.

The overall model fits are shown in Table 16. For EAM and HyGene, the fits shown are for the best fitting parameters for those models for each participant. The fits for the delayed session were then calculated by using the best fitting parameters from the immediate session for both EAM (threshold) and HyGene (encoding). However, the decay parameter was then also fit for HyGene. At the aggregate level, the best fitting model was the random model for both the immediate and delayed sessions.

Table 16
Aggregated Model Fits to Participant Cue Selection Pattern and Number of Cues Selected for Experiment 2

		Random	TTB	WADD	EAM	HyGene
Immediate	G^2	38033.5	92451.2	76092.1	44017.12	54123.07
	BIC		-54417.7	-38058.6	-5991.72	-16097.7
Delayed	G^2	37828.6	90676.4	77444.1	44996.02	49153.56
	BIC		-52847.8	-39615.5	-7159.32	-11333.1

Note: BIC is calculated in comparison to the random model.

Each participant was also classified based on which model had the best BIC compared to the random model as shown in Table 17. If the BIC was less than 0, then participant was classified as best fit by the random model. Again, the majority of participants were classified as using a random model.

Table 17
Proportion of Participants Classified as Best Fit by Each Model When Fitting Cue Selection Pattern and Number of Cues Selected

	Random	TTB	WADD	EAM	HyGene
Immediate	.68	.05	.05	.11	.12
Delayed	.53	.03	.05	.21	.18

Model fits were also conducted only on the pattern of cue selections without including fit to the number of cues selected. This was done because the patterns of selection shown in both Figure 12 and Figure 13 suggests that participants were not selecting cues in a truly random matter. It is possible that the number of cues participants were selecting could not be well accounted for with any of the models while the pattern of cue selections could be. It is important to note that results from this model fitting help determine the potential cause of the poor model fits above, but do not provide compelling evidence for any models because it ignores the number of cues selected for which all of these models make predictions.

Table 18 shows the model fits when the number of cues selected was not included in the fitting algorithm. When the number of cues was no longer being fit, both HyGene and EAM fit the data better than a random model, but the evidence accumulation model provided the best fit to the data for both the immediate and delayed sessions.

Table 18
Aggregated Model Fits to Participant Cue Selection Pattern Only for Experiment 2

		Random	TTB	WADD	EAM	HyGene
Immediate	G^2	27411.2	56002.1	27411.2	24197.4	26521.5
	BIC		-28590.9	0	3205.7	1164.6
Delayed	G^2	27206.3	5556.7	27206.3	23780.1	24702.4
	BIC		-28358.4	0	3418.1	2487.7

Note: BIC is calculated in comparison to the random model.

Participants were also classified based on which model had the best BIC compared to a random model as shown in Table 19. In the immediate session, more participants were classified as using EAM compared to the other models. In the delayed session, approximately equal numbers of participants were classified as using EAM and HyGene.

Table 19
Proportion of Participants Classified as Best Fit by Each Model When Fitting Cue Selection Pattern Only

	Random	TTB	WADD	EAM	HyGene
Immediate	-	.05	.21	.45	.29
Delayed	-	.05	.21	.38	.36

Note: WADD and random model make the same predictions

5.2.4 Discussion

Overall, this experiment does not provide strong evidence for a memory-based account of cue-based inferences. The goal of this experiment was to test the predictions of the accessibility-based framework in terms of the effects of dispersion of validity and retention interval on cue search behavior and decision outcome. Only one hypothesis based on the accessibility framework was confirmed. This experiment did find some, but not all, of the expected differences between the two validity dispersion conditions. The two conditions did not differ in the number of cues selected nor the amount of decision time. However, those in the low dispersion condition were more likely to select the compensatory option compared to those in the high dispersion condition. Further, there were differences in cue selection behavior between these conditions such that those in the low dispersion condition were more likely to select the fourth most valid cue and were more likely to select the third most valid cue when few cues were selected.

Taken all together, these results suggest that dispersion conditions differed in the way in which the cues were used to make decisions even though they did not differ in the amount of information selected nor decision time. This replicates prior work showing that participants' actual decisions match the ecological structure of the environment (Bröder, 2003; Lawrence et al., 2018a; Rieskamp & Hoffrage, 2008; Rieskamp & Otto, 2006) and replicates findings from Experiment 1. At the same time, it challenges work that suggests that participants search information differently under these environments because there were no differences in the number of cues selected or decision time. As noted above, the adaptive toolbox, EAM, and HyGene all assumed that there would be changes in the number of cues searched between the different dispersion conditions.

There was also no evidence for the expected differences between the immediate and delayed trials in the number of cues selected, cue preferences, or decision time. Opposite to what was expected, participants took longer on decisions in the immediate test trials compared to the delayed test trials. Which could be taken as evidence for a change in the way participants were making decisions after a delay, although not in the way that was expected. However, there were no differences between immediate and delayed test trials in which company was selected when compensatory and noncompensatory strategies differed. And the only difference in cue preferences was in the order of selection for the third and fourth most valid cues. This suggests that participants did not substantially change the way they searched for information or the way they used the information after the week delay. One potential account for the change in decision time is that participants were simply not trying as hard in the delayed test trials as evidenced by poorer performance compared to the immediate test trials.

One possible explanation for why this experiment failed to find many differences between conditions is the number of cues participants were selecting at test in both the immediate and delayed sessions. Participants tended to select more cues than necessary to maximize their final point total, especially in the high dispersion condition. Because most participants were selecting at least four cues, this resulted in a sort of ceiling effect for the comparison between the immediate and delayed sessions both in terms of the number of cues selected and preference for individual cues. This may also partially explain the lack of difference between the two dispersion conditions in terms of the number of cues and decision time. The cost of the cues was set in a way to try to encourage selective behavior but may have been set too low. A future study could specifically manipulate the cost of the cues to determine the effect that has on cue selection.

When fitting the number of cues selected and the pattern of cues selected, none of the models tested in the model fitting provided a good fit to the data. Again, this is likely because of the number of cues participants were selecting. Participants generally showed a preference for more valid cues over less valid cues, suggesting that their cue selection behavior was not entirely random. In fact, when the number of cues was not included in the model fitting procedure, both EAM and HyGene outperform the random model with EAM providing the best fit. Yet, the number of cues the participants selected was best accounted for by a random model. Both TTB and HyGene assume that few cues are selected, typically one or two for HyGene and a single cue for TTB. At low levels of the evidence threshold, EAM also assumes few cues are selected. However, at higher levels, EAM tends to favor a larger number of cues selected, typically three or five. This is likely why EAM provided the second-best fit to the data. Yet, it is important to note that EAM also assumes that cues are checked exclusively in order of validity. The model fitting procedure used assumed stochastic selection of cues for all models rather than deterministic; this was done to put the other models in the same terms as HyGene which does not assume a deterministic cue order. However, this means that the fitting procedure does not fully capture true model behavior for those models that assume a validity-based search order.

Although this experiment failed to provide evidence for a memory-based account for cue-based inferences, it also does not provide compelling evidence for any other account. The toolbox account assumes that participants follow a specific strategy during cue-based inference, but very few participants could be classified as using TTB or WADD. Moreover, the lack of differences in process data (decision time and the number of cues selected) between the dispersion conditions also challenges the idea of different strategies being used. The evidence accumulation account provided the best fit to the data but still failed to fit better than a random model when the number of cues selected was taken into account. As noted above, future research could specifically manipulate the cost of cue information to determine how that affects participant behavior as this likely influenced selection behavior in the current study.

5.3 Experiment 3: Cue-Based Inferences from Memory

The first two experiments tested predictions of the accessibility framework in environments in which the cue information was available to the participants. Yet, arguments around the adaptive toolbox often focus on the distinction between decisions made from givens and decisions made from memory. The above experiments are examples of decisions made from givens because at least some of the cue values were available to the participants while they were making their decisions. In contrast, participants must retrieve cue information from memory in decisions from memory. Although most of the studies testing different accounts of cue-based inferences use decisions from givens, Gigerenzer et al. (1999) argue that the tools within the adaptive toolbox, especially Take-the-Best, are more suited for decisions from memory. Moreover, they also argue that decisions from memory are more common than decisions from givens. Thus, showing that the accessibility framework can account for decisions in this context is important.

The primary reason for a lack of research on cue-based inferences from memory is the methodological challenge. In decisions from given information, testing assumptions about cue preferences and search behavior is rather straightforward because researchers can directly observe these behaviors. When decisions are made from memory, however, it is much more difficult to uncover the exact processes that participants use to make their decisions. Simply looking at final decisions does not provide compelling evidence for or against any particular frameworks. As noted above, the adaptive toolbox, single-strategy frameworks, and the accessibility framework can often result in the same decisions despite different processes.

Yet, some work has been done to test different accounts of cue-based inferences in decisions from memory. A few studies (Broder & Gaissmaier, 2007; Bröder & Schiffer, 2003) have used both reaction times and decision outcomes to classify participants as using different strategies. They found that reactions times corresponded well with what would be expected for the different strategies. In general, reaction times were longer the more cues that had to be checked before one discriminated, assuming people were searching in validity order. Work has also been done to show that parallel constraint satisfaction can account for decisions from memory by looking at reaction time, decision, and confidence (Glockner & Hodges, 2011). In this study, they found that PCS provided a better account of participant behavior than WADD, TTB, or an equal weights strategy.

However, these studies do not consider the implications of actual memory processes involved in decisions made from memory. This is likely because the frameworks being tested do not have memory mechanisms. In the studies discussed above, all cues were assumed to be equally memorable. But returning to Gigerenzer's (1999) argument that decisions from memory are more realistic, it is also likely more realistic that some cues are more accessible in memory than others. For example, some cues may be easier to retrieve because they have been seen more often than others. The other frameworks discussed in this paper cannot account for differences in cue accessibility if they do not match what would be expected based on cue validity. Note, it is possible that PCS could allow memory to influence subjective cue validity, but they do not formally specify the processes involved in determining subjective cue validity. In contrast, the accessibility framework, being a memory-based model, should operate well in decisions from memory.

The goal of this experiment was to test the accessibility framework in the context of decisions from memory. To achieve this goal, memory for specific cues was manipulated. In a previous experiment, the frequency of cue use during training was shown to affect cue preferences and later decisions (Lawrence et al., 2018a). However, in that experiment, cue values were provided to the participants. The current study sought to replicate these results in the context of inferences from memory. In the following experiment, participants learned the values of cues for the different options they would later be asked to compare. The frequency of learning for individual cues was manipulated such that one cue, either the most valid cue or the least valid cue, had more learning trials than the others. Note, however, that all cue values should have been learned but those with more training trials were likely to be more accessible. Like the previous experiments, the dispersion of cue validity was also manipulated. The different conditions were compared on decision time, decision outcome, and confidence.

The accessibility framework makes several predictions about the effects of these manipulations on cue-based inferences. Participants' decision times and decisions outcomes should be affected by the manipulation of cue validity such that decisions are slower and match a compensatory strategy in the condition with low validity dispersion compared to the one with high. These general predictions match the other frameworks. But the accessibility framework also predicts an effect of the memory manipulation, which the other frameworks cannot predict. When the best cue is trained most often, decision times should be faster in both the high dispersion and low dispersion conditions compared to when the worst cue is trained most often. This is because the manipulation boosts the accessibility of the best cue, making it more likely to be the only cue available, especially when cues are highly dispersed. The effect of the manipulations on confidence is more complex because it partially depends on the cue patterns for the options being compared (see simulation section for more detailed predictions). In general, the accessibility framework predicts that the manipulation of memory and dispersion of validity will affect confidence. Before describing the specifics of the experiment, a simulation of the HyGene model will be briefly discussed to demonstrate these predictions.

5.3.1 Simulation

Similar to Experiment 2, simulations were conducted to demonstrate the behavior of the HyGene accessibility model in this experiment. The model was trained on the different datasets used for training participants using the method described in the simulation section. The effect of the frequency manipulation was simulated by fixing the encoding parameter to 1 for the cue that was trained most frequently while the encoding of the other cues was set to .6. An alternative version of HyGene was also simulated based on the results from Experiment 2, which found that participants selected more cues than HyGene predicts. In the alternative version of HyGene (activation-only HyGene), all cues are available in memory and decisions are based made by weighing cues by their activations in memory and summing those values for each option. This model demonstrates the effect of differences in memory activation due to the manipulations without the cue selection processes being implemented. Simulation results focus on decision outcome, the number of cues generated, and confidence. Confidence in both versions of HyGene was determined by the support for the favored option over the combined support for both options, where support is based on the sum of cues available in working memory weighted by their accessibility. The discussion of the simulation will focus on the effect of the frequency manipulation because the effect of the dispersion manipulation has been discussed extensively in previous sections.

5.3.1.1 Decision Outcome

The proportion of options corresponding to a compensatory decision strategy for cases in compensatory and noncompensatory differed was calculated. This was done both for standard HyGene and activation-only HyGene. However, the pattern was the same for both models. In general, when the worst cue was trained most often, the compensatory option was more likely to be selected (see Figure 14). This effect was larger in the low validity dispersion condition compared to the high validity dispersion condition.

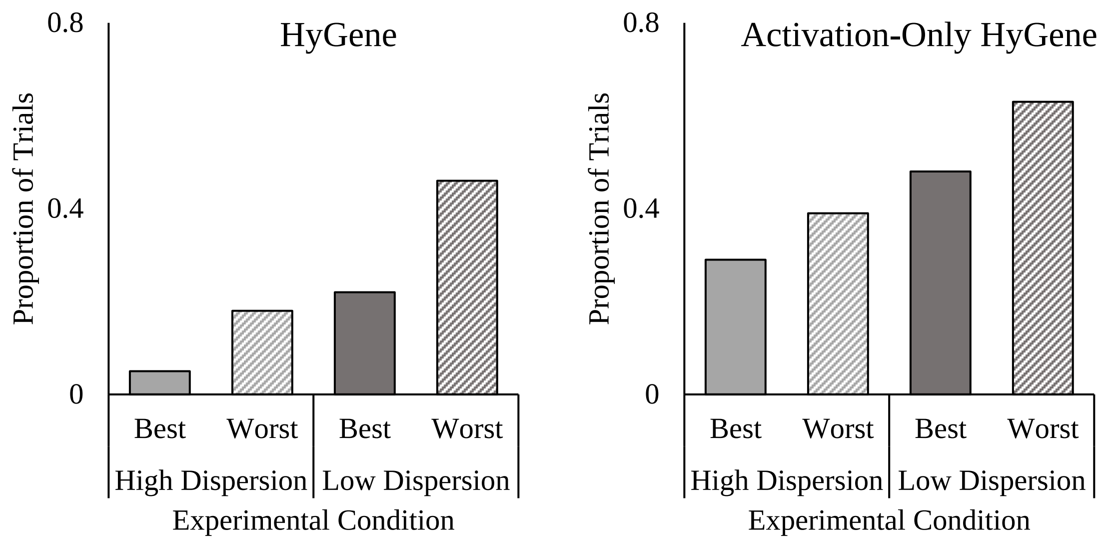

Figure 14. Proportion of decisions matching a compensatory rule when compensatory and noncompensatory differed by dispersion and frequency manipulations

5.3.1.2 Number of Cues Generated

As shown in Figure 15, the average number of cues generated was larger when the worst cue was trained most frequently compared to when the best cue was trained more frequently. Although HyGene does not make a direct decision time estimate, the number of cues generated was used as a proxy for decision time when comparing model performance and participant performance. This means the model assumes more time will be taken on decisions in which the worst cue was trained most frequently. Activation-only HyGene assumes all cues are always generated so it makes no assumptions about changes in decision time.

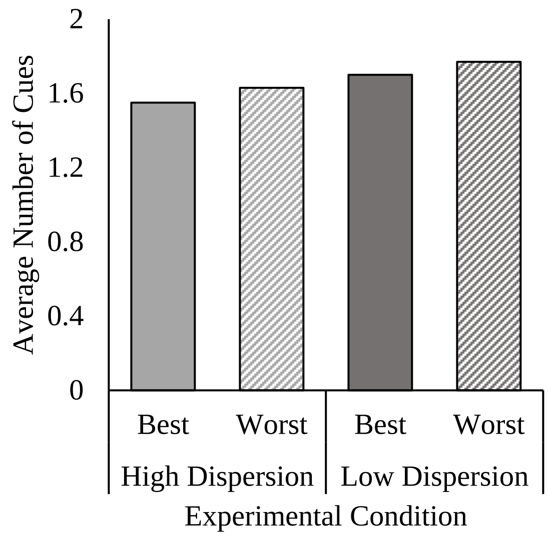

Figure 15. Average number of cues HyGene generated by experimental condition with an encoding parameter of .6 for all but most frequent cue.

5.3.1.3 Confidence

Average confidence for trials in which the best cue and the worst cue discriminated but compensatory and noncompensatory strategies should select the same option was compared across different simulated conditions. These trials were selected because they are the trials for which the conditions would be expected to differ most because the conditions should differ most in terms of accessibility of both the best and worst cue. Both versions of HyGene showed the same pattern. Confidence was lower in the low dispersion condition compared to the high dispersion condition, as shown in Figure 16. Confidence was also lower when the worst cue was trained most frequently compared to when the best cue was trained most frequently. This is because the increased accessibility of the less valid cues gave the unselected alternative a weighted value closer to the selected alternative.

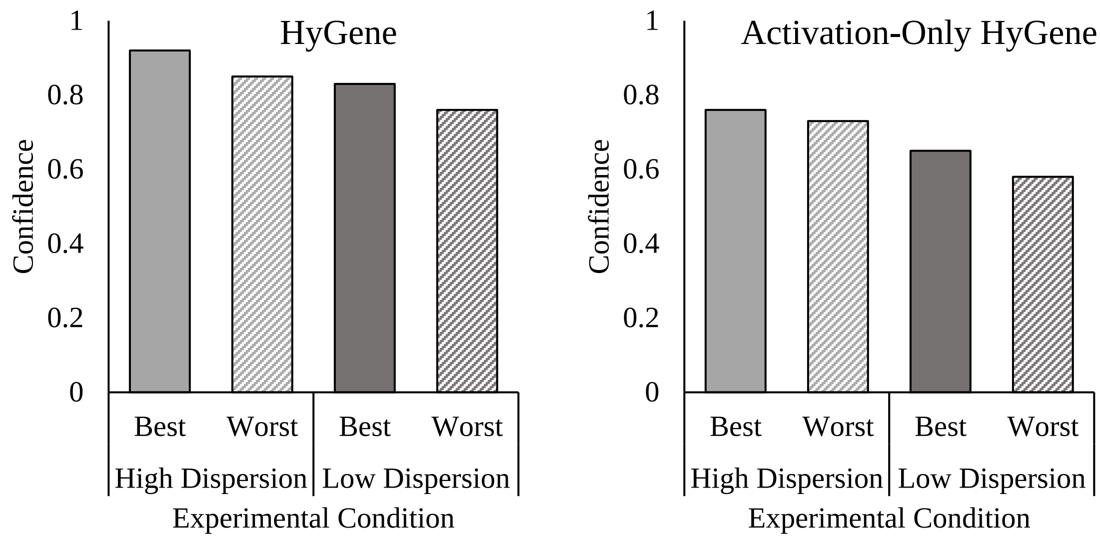

Figure 16. Average confidence of HyGene and activation-only HyGene by experimental condition with an encoding parameter of .6 for all but most frequent cue.

Overall, the simulations demonstrate that the accessibility framework makes predictions in regards to the effect of manipulations of individual cue's encoding quality on decision outcomes, the number of cues used, and confidence. When the worst cue is encoded the best, decisions are more likely to match compensatory strategies, more cues are used, and confidence tends to be lower. This effect interacts with the dispersion of validity, such that the effect is slightly larger when the dispersion is low. These predictions were tested in the following experiment.

5.3.2 Method

There were 186 participants in this study. However, participants that failed to reach a threshold of 80% correct in the test of recall for the learned cue values were excluded from data analyses. They were excluded because the goal of this experiment was to compare performance assuming that cue values were learned. Without this assumption in place, it is difficult to draw conclusions from the participant's data. This resulted in 119 participants being included in the study (30 in all but the high dispersion condition with the best cue trained most frequently, which had 29). Participants were recruited for this study using the online experiment management system at Georgia Institute of Technology. Participants received course credit for their participation.

The design of this experiment was a 2 (number of learning trials) by 2 (dispersion of cue validity) with both variables manipulated between participants. The number of learning trials was manipulated by changing which cue had the most learning trials: the most valid cue or the least valid cue. There were also two conditions for the dispersion of cue validity: a standard deviation of .04 (low) or .13 (high).

Unlike the previous studies, this study was presented as a decision between which fictional planets were more likely to support life. Each planet was described by four cues: geochemistry, microenvironments, orbital eccentricity, atmosphere. These cues were either positive or negative. The validity of these cues was manipulated so that the cue validities were either highly dispersed or not, as shown in Table 20. The labels and the spatial locations of the cues were random between participants, meaning the cues listed in Table 20 did not appear in the same location or have the same label for all participants.

Table 20
Validity of the Cues by Condition for Experiment 3

	High Dispersion	Low Dispersion
Cue 1	.88	.75
Cue 2	.76	.72
Cue 3	.64	.68
Cue 4	.52	.66
Mean	.70	.70
SD	.13	.04

This study consisted of three phases: cue learning, validity learning, and test. During the cue learning phase, participants were told that they needed to learn four distinct characteristics of eight different planets. This phase was followed by a four-part procedure for cue value learning similar to the one used by Glöckner & Hodges (2011) and shown in Figure 17. First, participants were shown each planet with all four cues. Second, they were shown individual cue values for all cues for all planets in random order. Third, they were asked to recall individual cue values for all cues for all planets in random order and they were provided with feedback. It was during this part of cue learning that the manipulation took place. Participants saw each cue twice for each planet except for the cue that was being manipulated, which they saw four times. Fourth, they were presented with each planet and asked to recall all four cues for each, and then they were provided with feedback. The third and fourth parts of the cue learning phase were repeated three times.

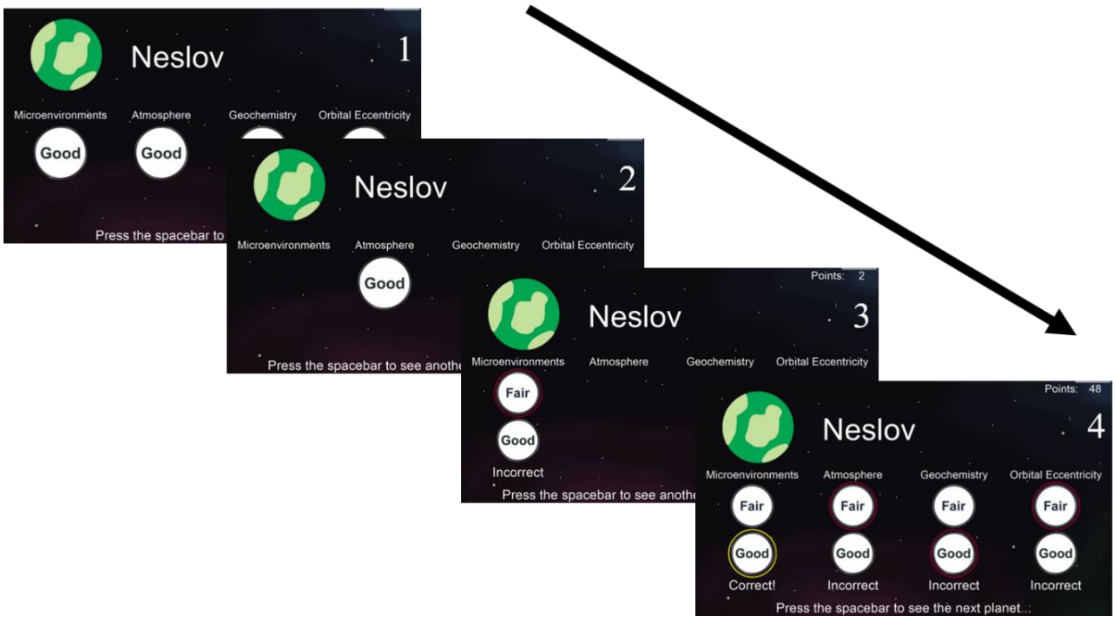

Figure 17. Example of trials in cue learning phase. Each image corresponds to one of the four phases of the cue learning phase.

Once the cue learning phase was complete, participants completed a validity learning phase. During the validity learning phase, participants completed 100 paired comparisons in which they needed to decide which planet was more likely to support life. Participants were instructed that they needed to learn which cues were better at predicting which planet was more likely to support life. Like the validity learning phase in Experiments 1 and 2, all four cues were presented during this phase and participants were provided with feedback. The planets during this phase were not the same eight planets they learned the cues for previously. However, they were instructed that the relationship between the cues and the criterion is the same for these new planets as the planets they learned previously.

At the end of the validity learning phase, participants completed a refresher on the cue values for the cues for the eight planets they learned previously and a final memory check. The refresher consisted of repeating the third and fourth parts of the cue learning phase once more. The fourth part was then repeated a second time without feedback to get a measure of learning. Then they moved on to the test phase as shown in Figure 18. During this phase, they were instructed to use what they had learned in the previous phase to decide which of the planets they learned about were more likely to support life. During this phase, they were presented with all possible combinations of the eight planets they learned. However, the cues values for these planets were not presented on the screen but instead needed to be retrieved from memory. The decision outcome, decision time, and confidence were measured for each decision.

Figure 18. Example of test phase portion for Experiment 3.

5.3.3 Results

The results of this experiment are organized as follows. First, learning checks are discussed. Learning was checked in terms of recall of cue values, validity learning, and performance on test trials. Then the results for critical trials are discussed. These are the trials for which both the worst cue and the best cue discriminated but compensatory and noncompensatory decision strategies select the same option. Third, the results for test trials in which compensatory and noncompensatory strategies make different selections are discussed. The results are separated by trial type because the predictions discussed above, particularly for planet selected and confidence, were specific to certain types of trials. Finally, participant performance is compared to model predictions.

5.3.3.1 Learning Checks

As noted above, only participants who exceeded a threshold of 80% correct on the memory check for cue learning were included in the analyses. The average proportion correct of included participants was .957, while the average of excluded participants was .626. A generalized linear model with a binomial logit link was used to test whether dispersion of validity or cue frequency affected recall of the cue values for which the frequency of training was manipulated. For the most valid cue, dispersion of validity (χ^2 (1) = 1.11, p = .29) and the frequency manipulation (χ^2 (1) = .40, p = .53) did not significantly affect recall and there was no interaction (χ^2 (1) = .07, p = .79). For the least valid cue, the frequency manipulation did significantly affect recall (χ^2 (1) = 4.72, p = .03) such that participants for which the worst cue was trained most frequently were more likely to recall the cue value (O = 3.64) compared to those with the best cue trained most frequently (O = 2.78). However, the proportion of those recalling the cue value for the least valid cue was .94 in the condition with the best cue trained most frequently and .97 in the condition with the worst cue trained most frequently so it is unlikely this difference substantially affected performance at test. Dispersion condition (χ^2 (1) = 0.06, p = .81) did not significantly affect recall of the least valid cue and there was no interaction (χ^2 (1) = 0.02, p = .88).

Validity learning was checked both over the validity learning trials and at the final test trails. For the validity learning trials, a generalized linear model with a binomial logit link was used to test if trial, dispersion of validity, and cue frequency affected performance during the training trials. Trial (χ^2 (1) = 7.24, p = .007) significantly affected performance during training such that performance increased across training trials, suggesting that participants were learning. Dispersion of validity also significantly affect performance during training (χ^2 (1) = 11.76, p < .001), such that those in the low dispersion condition (85.38% correct) performed better than those in the high validity condition (82.83% correct). The frequency manipulation did not significantly affect performance (χ^2 (1) = 1.04, p = .31) and there were no significant interactions. A binomial test, collapsed across conditions, also showed that the proportion of correct responses on the last quarter of training trials (.85) was above a chance level of .5 (z = 38.19, p <.0001), also suggesting that participants were able to learn.

For test trials, a generalized linear model with a binomial logit link was also used to test the effect of dispersion of validity and the frequency manipulation on participant's performance. This analysis was conducted on trials in which compensatory and noncompensatory strategies resulted in the same planet being selected. There were no significant effects of dispersion of validity (χ^2 (1) = .01, p = .92) or the frequency manipulation on participant's performance (χ^2 (1) = .44, p = .51). On average, participants selected the correct planet on approximately 82% of the trials, which is consistent with the learning checks from the validity learning phase. In general, participants seemed to learn how to use the cues correctly.

5.3.3.2 Results for Critical Trials

The following analyses were conducted on trials in which both the best cue and the worst cue discriminated but compensatory and noncompensatory strategies should result in the same option being selected. As described above, these trials were selected because they were the trials for which the manipulations were expected to have the strongest result. Hypotheses for these trials focused on confidence, but option selected and decision time were also analyzed.

To examine the effects of the manipulations on confidence, a generalized linear model with an identity link was also used to test the effect of dispersion of validity and the frequency manipulation on participant's decision confidence. Two models were run; one for correct responses and one for incorrect responses. For correct responses, decision confidence differed significantly between dispersion conditions (χ^2 (1) = 5.12, p = .02), such that participants were more confident in the high dispersion condition (M = 70.23, SD = 25.37) compared to the low dispersion condition (M = 61.71, SD = 25.83). The frequency manipulation did not significantly affect decision confidence (χ^2 (1) = .44, p = .51) nor was there a significant interaction (χ^2 (1) = .28, p = .60). For incorrect responses, there was no effect of dispersion (χ^2 (1) = 1.09, p = .30) or frequency (χ^2 (1) = 2.11, p = .15) and there was no interaction (χ^2 (1) = .07, p = .79). Average confidence for incorrect responses was 54.26 (SD = 27.96).

Further analyses were conducted to check for differences in bias and calibration. Bias was calculated as the difference between overall performance and average confidence for each participant. Calibration was calculated as the mean square deviation from confidence to percent correct weighted by bin size for each participant. For this calculation, confidence judgments were separated into five bins. A general linear model was used to test the effect of dispersion of validity and frequency manipulation on bias. There was no effect of dispersion ($F(1,115) = 1.06$, $p = .31$), or frequency ($F(1,115) = 0.25$, $p = .62$), and there was no interaction ($F(1,115) = 0.16$, $p = .69$). Average bias was -.04 (SD = .27), suggesting slight overconfidence. A general linear model was used to test the effect of dispersion of validity and frequency manipulation on calibration. There was a significant effect of dispersion of validity ($F(1,115) = 4.39$, $p = .04$), such that those in the high dispersion condition (M = .13, SD = .11) were better calibrated than those in the low dispersion condition (M = .18, SD = .14). There was no effect of frequency ($F(1,115) = 0.89$, $p = .35$) and there was no interaction ($F(1,115) = 1.04$, $p = .31$). Figure 19 shows calibration between conditions at the aggregate level, with the condition in which the best cue was trained most frequency and dispersion was high being the most well calibrated.

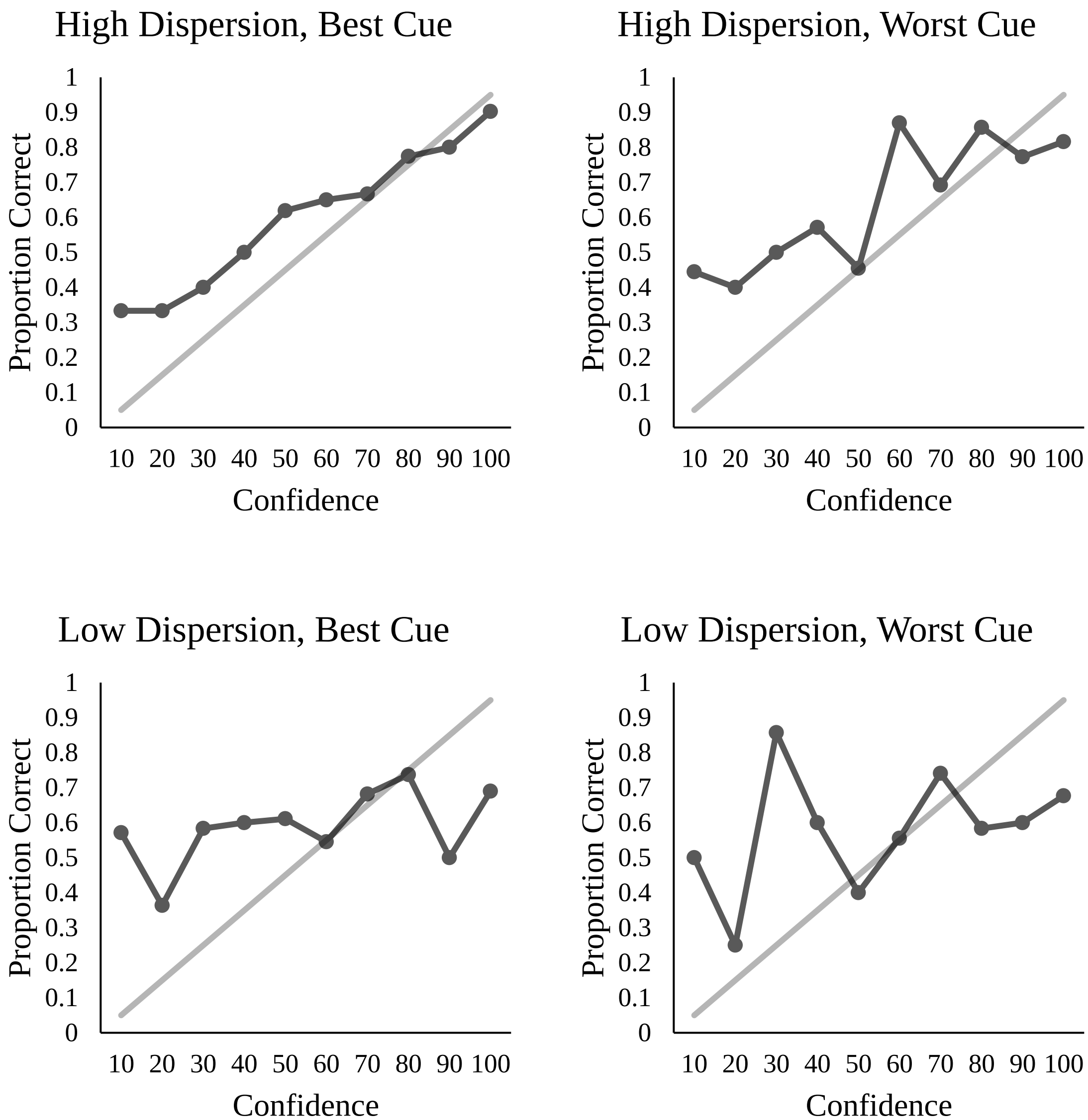

Figure 19. Calibration curves by dispersion and frequency conditions.

To determine whether the conditions performed differently on the critical trials, a generalized linear model with a binomial logit link was also used to test the effect of dispersion of validity and the frequency manipulation on participant's decisions. Decisions differed significantly between dispersion conditions (χ^2 (1) = 4.91, p = .03), such that the odds of selecting the correct option was higher when dispersion was high (O = .89; 71%) compared to when dispersion was low (O = .44; 61%). The frequency manipulation did not significantly affect decisions (χ^2 (1) = .17, p = .68) nor was there a significant interaction (χ^2 (1) = .09, p = .77).

Finally, the effect of the manipulations on decision time for critical trials was also tested. A generalized linear model with an identity link was also used to test the effect of dispersion of validity and the frequency manipulation on the participant's decision time. Decision time was log transformed to better handle outliers. There were no significant differences of dispersion of validity (χ^2 (1) = .58, p = .45) or the frequency manipulation (χ^2 (1) = .18, p = .67) and there was no interaction (χ^2 (1) = .30, p = .59).

5.3.3.3 Results of Trials in which Compensatory and Noncompensatory Strategies Differ

The following analyses were conducted on trials in which compensatory and noncompensatory strategies differed in which planet should be selected. Hypotheses for these trials focused on differences between conditions on decision time and option selected. However, additional analyses of confidence were also conducted to further elucidate differences found between conditions.

To determine whether the manipulations affected apparent decision strategy, a generalized linear model with a binomial logit link was also used to test the effect of dispersion of validity and the frequency manipulation on participant's decisions.

Decisions differed significantly between frequency conditions (χ^2 (1) = 4.77, p = .03), such that the odds of selecting the compensatory option was higher when the worst cue was trained most frequently (O = 2.56, 92.54%) compared to when the best cue was trained most frequently (O = 1.69, 84.33%). Dispersion of validity did not significantly affect decisions (χ^2 (1) = 1.53, p = .22) nor was there a significant interaction (χ^2 (1) = .28, p = .59).

A generalized linear model with an identity link was also used to test the effect of dispersion of validity and the frequency manipulation on participant's decision time. Decision time was log transformed to better handle outliers. Log decision times differed significantly between dispersion conditions (χ^2 (1) = 5.04, p = .02), such that participants took more time making decisions in the high dispersion condition (*M* = 5402.04, *SE* = 360.90) compared to the low dispersion condition (*M* = 4081.02, *SE* = 319.35). The frequency manipulation did not significantly affect decision time (χ^2 (1) = 1.37, p = .24) nor was there a significant interaction (χ^2 (1) = .65, p = .42).

Finally, a generalized linear model with an identity link was also used to test the effect of dispersion of validity and the frequency manipulation on participant's decision confidence for trials in which the compensatory and noncompensatory strategies differed in the option selected. Two models were run, one for when the compensatory option was selected and one for when the noncompensatory option was selected. For responses matching a compensatory strategy, decision confidence differed significantly between dispersion conditions (χ^2 (1) = 5.12, p = .02), such that participants were more confident in the low dispersion condition (M = 88.28, SD = 19.38) compared to the high dispersion condition (M = 78.53, SD = 22.60). The frequency manipulation did not significantly affect decision confidence (χ^2 (1) = .97, p = .33) nor was there a significant interaction (χ^2 (1) = .02, p = .89). For responses matching a noncompensatory strategy, there was no effect of dispersion (χ^2 (1) = 3.55, p = .06) or frequency (χ^2 (1) = 2.66, p = .10) and there was no interaction (χ^2 (1) = .17, p = .68). Average confidence for noncompensatory responses was 49.48 (SD = 28.15). It is worth noting that the sample size for the model for noncompensatory decisions was much smaller than for compensatory because all but one participant made at least one compensatory response but only 37 participants made at least one noncompensatory response.

5.3.3.4 Comparison of Participant Data to Model Simulations

Although formal model fitting was not conducted for this experiment, participant performance was compared to predictions from TTB, WADD, EAM, and both versions of the HyGene accessibility framework. Predictions for all but the HyGene model were determined by running each model through each test case once. Because HyGene is a stochastic model, predictions were determined by averaging simulation performance across 500 simulation runs for each test case. A HyGene model in which decision behavior was determined only by normalized activation in memory was also included in the model comparisons. This model assumes all information is available in working memory, and the information is weighted by memory activation to arrive at a decision. This model was included to determine whether HyGene memory activations can predict participant performance. The previous study found that HyGene selected less information than participants so including this model should help elucidate the effect of HyGene cue selection behavior on model performance.

First, overall adherence rates of participants' planet selections for each case were compared to model predictions as shown in Table 21. No model consistently outperformed the other models. However, TTB and standard HyGene never fit best. Activation only HyGene fit the best or tied for the best fit in three conditions. EAM and WADD provided the best fit or tied for the best fit in two conditions.

Table 21

Adherence Rates for Participants' Decision Outcomes Compared to Model Predictions

Dispersion	Frequency	TTB	WADD	EAM	HyGene	Activation Only HyGene
High	Best	0.7	0.7	0.7	0.7	0.73
	Worst	0.69	0.68	0.69	0.78	0.83
Low	Best	0.69	0.81	0.81	0.69	0.76
	Worst	0.7	0.86	0.86	0.8	0.86

Second, model predictions for decision time were compared to participant performance. Decision time for the models was simply the number of cues the models checked before making a decision for each test case. Model predictions were correlated with the average amount of time participants took on each individual case by experimental condition. Model decision time and average participant decision time were correlated for all models that made variable predictions for the number of cues checked as shown in Table 22. Models that assumed all cues were checked on every trial (WADD, EAM at high thresholds, and activation only HyGene) could not be fit. HyGene decision time estimates were most strongly correlated with participant decision time for all conditions, although in the high dispersion/ best cue condition this was a strong negative correlation. Also note, that HyGene did not vary much in terms of the number of cues selected between cases within each condition because the number of cues was just based on availability in memory which was not case dependent. This could mean that HyGene provided the best correlation simply because it did not make extreme predictions like the other models.

Table 22
Correlation between Participant Decision Times and Models' Conceptual Decision Times

Dispersion	Frequency	TTB	EAM	HyGene
High	Best	-.19	.28	-.54
	Worst	-.12	-.11	.36
Low	Best	-.19	-.19	.36
	Worst	-.24	-.24	.33

Note: Estimates for EAM and HyGene reflect parameters with the strongest correlation

Finally, model predictions for confidence were also compared to participant performance. Confidence in TTB was calculated as the validity of the cue upon which the decision was made (as described in Glöckner & Broder, 2011). Confidence in WADD and EAM were calculated as the odds of selecting the planet over the combined odds of selecting either planet. For HyGene, confidence was calculated by weighing each cue by its normalized activation and then summing those values for each planet. The weighted sum for the planet selected was then divided by combined the weighted sum for both planets. Again, no model consistently correlated best with participant performance, as shown in Table 23. However, standard HyGene never provided the highest correlation with participant data and TTB only provided the strongest correlation in one condition. Activation-only HyGene and EAM were both relatively strongly correlated with participant performance for all conditions, but activation-only HyGene was the only model that was consistently positively correlated.

Table 23

Correlation between Participant Confidence and Model-Predicted Confidence

Dispersion	Frequency	TTB	WADD	EAM	HyGene	Activation Only HyGene
High	Best	-0.22	0.19	-0.24	-0.22	0.26
	Worst	-0.33	-0.21	-0.33	-0.08	0.22
Low	Best	-0.13	0.63	0.63	-0.21	0.35
	Worst	-0.27	0.55	0.55	-0.18	0.65

Note: Estimates for EAM and HyGene reflect parameters with the strongest correlation

5.3.4 Discussion

This experiment provides partial support for a memory-based account of cue-based inferences from memory. The frequency manipulation affected which planet was selected at test when compensatory and noncompensatory options differed. As hypothesized, those in the condition in which the worst cue was trained most frequently were more likely to select the compensatory option compared to those in the condition in which the best cue was trained most frequently. This extends previous research in which frequency of training was found to affect apparent decision strategy in cue-based inferences from given information (Lawrence et al., 2018a). Moreover, none of the current models for cue-based inferences, other than the accessibility-based framework, can account for the effect of training frequency on decisions. Although differences in recall for the least valid cue were found for the different frequency conditions, the differences were slight and cannot account for the effect.

Despite evidence for an effect of the frequency manipulation on decision outcomes, there were no significant differences between the frequency conditions for decision time or confidence. It was hypothesized that those in the condition in which the worst cue was trained most frequently would take more time and show lower confidence on critical trials than those for which the best cue was trained most frequently. A lack of difference in decision time with the presence of a difference in decision outcome challenges some assumptions present in the literature. Assuming participants were using qualitatively different strategies, it challenges the supposition that compensatory strategies take longer than noncompensatory strategies. It also further challenges the assumption that decision outcomes reliably indicate actual decision strategy (see Dummel et al., 2016; Glöckner, Hilbig, & Jekel, 2014). It could be the case that participants were adjusting the weights of cues, resulting in apparently difference decisions strategies based on decision outcomes while actually using the same number of cues. However, based on model simulations, changing the way the cues were weighted should have resulted in a change in confidence, which was not found.

This experiment does not provide strong support for hypothesized effects of the dispersion of validity in decisions made from memory. Looking first at trials in which compensatory and noncompensatory strategies should select different planets, it was hypothesized that those in the low dispersion condition would be more likely to select the compensatory option compared to those in the high dispersion condition. However, there were no significant differences in decision outcomes between those in the low and high dispersion conditions for these trials. This challenges previous research on the effects of the dispersion of validity (Bröder & Schiffer, 2006) and the previous two experiments. Moreover, those in the high dispersion condition took longer on these test trials, opposite of predictions. They also showed lower confidence in these test trials when they selected the compensatory option. It is almost as though a compensatory strategy was the default strategy and when the ecology favored a noncompensatory strategy, participants took more time and were less confident in their decisions but still selected the compensatory option (Bröder, 2003; Glöckner & Betsch, 2012).

For trials in which both the most valid and least valid cues discriminated, no differences were specifically hypothesized for the dispersion conditions. Nevertheless, there were significant differences between dispersion conditions in terms of confidence and correct responses. For these trials, those in the low dispersion condition were less confident and less accurate. Moreover, their confidence was also more poorly calibrated with actual performance than those in the high dispersion condition. Although this was not a focal hypothesis, the lower confidence for the low dispersion condition matches predictions from the activation-based framework shown in Figure 16. Moreover, a post-hoc investigation of the simulation results also found poorer performance in the low dispersion condition (82% correct) compared to the high dispersion condition (96%). In some simulation trials, the less valid cues were more activated than the most valid cues which resulted in poorer performance in the low dispersion condition, especially when the worst cue was trained most often. This may explain what is happening with the participants in these trials.

The above results provide unexpected support for the accessibility-framework. None of the other models predicted poorer performance under the low dispersion condition for the trials discussed above. In fact, all of the other models predicted perfect performance on those trials because they assume cues are always weighted by validity. Yet, in terms of direct comparison to participant data, no single model provided the best matched to participant performance for all variables in all conditions. However, models that favored the use of few cues (TTB and HyGene) tended to not fit as well as those that favored most or all cues being used (WADD, EAM, and activation-only HyGene).

Considering both the results of the statistical tests and the results of the comparison of participant data to model predictions, it seems that participants tended to use most, if not all, cues when making decisions. This is consistent with the fact that participants tended to select the compensatory option over the noncompensatory option during test trials. This is also consistent with the results from Experiment 2, which found that participants tended to select most cues. However, this finding is a little surprising given that these decisions were made from information stored in memory, decisions for which TTB was specifically designed (Gigerenzer et al., 1999). Nevertheless, this consistent with previous research (Glöckner & Hodges, 2011) has found some evidence that people may continue to use compensatory strategies when decisions are made from memory.

Although the results of this experiment were not conclusive in terms of which framework provides the best account of cue-based inference, they do suggest some frameworks may be better than others. It seems particularly unlikely that participants were using an adaptive toolbox. Hardly any participants consistently selected the noncompensatory option when compensatory and noncompensatory strategies differed. Under an adaptive toolbox, participants in the high dispersion condition should have used a noncompensatory strategy, especially because these were decisions from memory. None of the results directly rule out the evidence accumulation model, but there were some results that only an accessibility-based framework could account for, such as the effect of the frequency manipulation. Yet, not all of the predicted results from the accessibility-based framework were confirmed.

One possible explanation for the lack of large differences between conditions is that a large number of participants had to be excluded from the analyses. Participants that could not recall at least 80% of the cue values were not included in the analyses. Because they could not recall cue values, their responses at test would have been difficult to interpret. It would be impossible to determine what information they were using to make their decisions. However, it is possible that the participants who could recall cue information may not be representative of the general population. Those who were able to recall all cue information may also have been more likely to consider all information at test. Although the cue learning phase was based on previous research (Glöckner &

Hodges, 2011), a future study could provide more test trials or better incentives for the participants to learn the cue values to get a larger proportion of participants above the 80% cutoff.

In general, the results of this study challenge assumptions found in prior research. It challenges the assumption that a majority of participants use a noncompensatory strategy when the dispersion is high (Bröder & Schiffer, 2006; Newell & Bröder, 2008) or when decisions are made from memory (Bröder & Schiffer, 2003; Gigerenzer et al., 1999). Moreover, much of the research on decision strategies assume that compensatory strategies take more time than noncompensatory strategies (see Glöckner et al. (2014) for an alternative account). However, in some trials, this study found a difference in decision time with no corresponding difference in decision outcomes and vice-versa. This could mean that compensatory strategies do not necessarily take longer, as argued by Glöckner and colleagues. Finally, the effect of the frequency manipulation on the option selected and the effect of dispersion on accuracy and confidence provides some evidence for the role of memory in cue-based inferences that cannot be accounted for by validity alone. Thus, challenging validity-based models and suggesting that accessibility-based models may be worth investigating further.

CHAPTER 6: GENERAL DISCUSSION

Since the work of Payne et al. (1988), researchers have been trying to understand how decision makers adapt to different decision environments. The primary goal of this dissertation was to test an accessibility account of adaptive decision-making in cue-based inferences. The accessibility framework was proposed to address the shortcomings of current frameworks for cue-based inferences. By basing cue use on memory accessibility, the accessibility framework addresses criticisms of current single strategy frameworks (evidence accumulation model and parallel constraint satisfaction) and the adaptive toolbox. Theoretically, the accessibility framework resolves problems with psychological plausibility, falsifiability, reliance on cue-validity, and inability to account for memorial effects found in the other frameworks. In order to test the predictions of the accessibility framework against predictions of current frameworks for cue-based inferences, three studies were conducted that manipulated both cue accessibility and cue validity.

In general, the studies found partial support for the role of memory in cue-based inferences. In two of the three studies, there was evidence for an effect of memory manipulations on decision outcomes. In the first experiment, participants preferred the compensatory option when primacy was manipulated in the less valid cue pair. This interacted with the manipulations of dispersion of validity such that the effect disappeared when the dispersion was low and the mean was high. In the third experiment, the frequency with which participants recalled cue values during training affected decision outcomes such that participants were more likely to select the compensatory option when the less valid was trained most frequently. Not only do these findings extend previous research that has shown participant's decision outcomes are sensitive to memory manipulations (Lawrence et al., 2018a; Platzer et al., 2014), they also challenge current frameworks of cue-based inferences. Because the current frameworks are validity-based, they cannot account for the effect of memory.

However, the studies failed to confirm all hypothesized effects of the memory manipulations. In the first experiment, there was only evidence for the preference of the primacy cue over a similarly valid cue in one condition. Moreover, there was evidence for a preference for the cue that discriminated correctly most often in the middle of each block rather than the beginning or end in several conditions. In the second experiment, there was no evidence for the hypothesized differences between the immediate test and the delayed test in terms of cue preferences, the number of cues selected, or decision outcomes. In fact, participants tended to take less time on the delayed decisions compared to the immediate decisions, which is opposite of what was hypothesized. In the third experiment, there were no differences between frequency conditions in terms of decision time or confidence.

Although these results challenge the accessibility framework, most of them are also not well explained within the other frameworks. Any effects of manipulations of memory, such as serial position effects and training frequency, cannot be accounted for with validity-based models (adaptive toolbox, EAM, and PCS) because they assume that participants' cue preferences should be based on validity alone. For example, participants should not show a preference between cues with the same validity no matter the serial position of correct discriminations at training. Any evidence for a strong preference between cues with the same validity, even if it is a preference for the cue that was correct most often in the middle of each block, challenges validity-based models.

In addition to providing partial support to hypotheses based on the accessibility framework in terms of the effects of memory manipulations, there was also partial support for the hypothesized effects of the validity manipulations. Specifically, there was evidence for a preference for the compensatory option when dispersion was low in both the first and second experiment, matching prior research that has found effects of dispersion manipulations (Bröder & Schiffer, 2006; Newell & Bröder, 2008). Because all frameworks make similar predictions for the effect of validity manipulations on decision outcomes, evidence for these effects does not provide as strong support for accessibility framework as evidence for the effects of direct memory manipulations. As noted above, all frameworks predict changes in decisions based on the decision environment but differ in the mechanism of those changes. However, in Experiment 3, dispersion of validity also affected confidence and accuracy for decision trials in which the most valid and least valid cues discriminated. Only the accessibility framework can account for differences in performance for these trials because the other frameworks are based on validity.

Again, there were also results related to the effect of the dispersion manipulations that could not be accounted for with the accessibility framework or the other frameworks. In the second experiment, there was no evidence of differences in the number of cues selected or decision time between the two dispersion conditions. There was also only evidence for differences in cue preferences for the third and fourth most valid cues and not the most valid cue. These results challenge the assumption that dispersion of validity affects cue selection and decision time, an assumption held by all frameworks except

PCS. Finally, there was also no evidence of a difference between dispersion conditions in terms of decision outcome for decisions in which compensatory and noncompensatory options should differ in the third experiment, challenging previous research (Bröder & Schiffer, 2006) and the previous two experiments.

The effect, or lack of effect, of the manipulation of dispersion in both Experiment 2 and Experiment 3 are particularly challenging to the adaptive toolbox and the evidence accumulation model. EAM and the adaptive toolbox assume more cues are selected and more time should be taken for decisions matching a compensatory strategy. In Experiment 2, there was an effect of dispersion on decision outcome but not on the number of cues selected. In contrast, those in the high dispersion condition in Experiment 3 took longer on test trials, despite no differences in decision outcomes. Neither of these findings is consistent with EAM or the adaptive toolbox. Although these results are also not consistent with predictions based on simulations of the accessibility model, they are not impossible under the theoretical framework. Within the accessibility framework, cues are not weighted by validity so it is possible for different instances of the model to select different options without differing in the number of cues generated due to differences in accessibility. For example, if a less valid cue is very accessible instead of the most valid cue, then the accessibility framework may choose the compensatory option without differences in the number of cues generated.

Another challenge to the adaptive toolbox is the lack of evidence for noncompensatory strategies in all experiments. Although some of the studies found differences in the proportion of participants selecting compensatory options, none of the experiments reported in this dissertation found that a majority of participants selected the noncompensatory option over the compensatory one when those strategies should differ. This is consistent with the literature which has found little evidence for TTB when validity is learned in training (Newell & Shanks, 2003; Newell et al., 2003; Rakow et al., 2005). However, the decision environment of the last experiment, especially in the high dispersion condition, should have resulted in a number of participants using TTB under the assumptions of the adaptive toolbox (Marewski & Schooler, 2011). Yet, very few participants selected noncompensatory options in this experiment.

In general, the results of the experiments support the assumption that compensatory strategies are the default strategy (Bröder, 2003; Glöckner & Betsch, 2008a; Rieskamp & Otto, 2006; Söllner et al., 2014). Most participants selected the compensatory option in all conditions. Moreover, in both Experiments 2 and 3, participants seemed to be using most of the cues. Glöckner and Hodges (2011) argued that noncompensatory strategies, TTB specifically, are only used when retrieval of information is effortful, such as in laboratory studies in which participants must access unfamiliar cue values and validities. It is possible that there are certain conditions under which participants may have used noncompensatory strategies, but that was not observed.

One potential explanation for the lack of evidence for Take-the-Best is that the cues in the experiments in this dissertation were uncorrelated. This was done in order to better control the relationship between validity and direct memory manipulations. However, there is evidence that people are more likely to use simple strategies, like TTB, when cues are intercorrelated (Dieckmann & Rieskamp, 2007; Lee & Zhang, 2012). Future research might test whether the conclusions from the studies in this dissertation hold when cues are correlated with each other. Although not simulated in this dissertation, the accessibility framework would likely make different predictions under correlated cues compared to uncorrelated cues. Extending the findings in this dissertation to environments with correlated cues would also be important for better understanding cue-based inferences outside of the laboratory because information is often redundant in natural environments (Gigerenzer & Brighton, 2009).

Nevertheless, evidence for a trend toward compensatory strategies supports the parallel constraint satisfaction model. This model was absent from the model comparisons in the experiments for this dissertation because many aspects of the model are not well specified, such as the search process and how cue validities are initialized. It is important to note, however, that it could account for many of the findings in this dissertation. Theoretically, it can account for the evidence that participants almost exclusively selected the compensatory option in many cases. It could also potentially account for the finding of no difference in decision time even when there were differences in decision outcome and vice versa because decision time is based on the maximization process rather than the number of cues searched. One could even argue that the effects of the memory manipulations could be accounted for by adjusting the initial cue validities to reflect memory accessibility.

However, the way PCS has been implemented results in more of a descriptive model rather than a process model. Many of its parameters cannot be understood psychologically within the consistency maximization process. As noted above, the process for determining initial cue validities is also not specified. In practice, the initial cue validities fed into the model are the ecological cue validities, but Glöckner and Betsch (2008a) make it clear that these could also be subjective cue validities.

Unfortunately, the process for determining initial cue validities is not well specified within in the model so it would be difficult for the model to account for memory effects in practice. Because of the shortcomings of this model, it is hard to endorse it as the best framework of cue-based inferences.

Although the experiments in this dissertation also do not provide conclusive evidence in support of the accessibility framework, they indicate how the framework could be improved. Both Experiments 2 and 3 showed that participants often used most of the cues when they were allowed to search for themselves. These findings are not consistent with predictions of the accessibility framework. The accessibility framework predicts different numbers of cues being used under different decision environments, and it very rarely uses more than three cues. This is because as soon as the most accessible cue is retrieved, no other cues can be retrieved. However, this process may need to be reconsidered. When an activation-only model, a model in which decisions were based on all cues weighted by memory accessibility, was compared to participant performance, it tended to provide a better match to participant data compared to the accessibility model with cue selection operating. This supports the conclusion that the accessibility framework's search process is limiting the model's performance.

One way to address this in the accessibility-based model is to allow it to make decision specific predictions for the cue generation process. Both PCS and EAM can handle decision specific predictions in terms of the number of cues used (EAM only), decision time, and confidence. However, the accessibility model only makes decision specific predictions for confidence. As is, the model retrieves all cues that will be used in the decision process based on activation alone and not based on the specific decision being made. For example, it does not retrieve more cues even if none of the available cues discriminate. Based on the experiment results, it is very unlikely that participants stopped searching cues when no cues discriminate. This could be adjusted by allowing the model to set the activation threshold back to zero after the maximum number of retrieval failures was reached but no discriminating cues were generated.

Because the accessibility-based framework is a decision model, decision specific predictions could also be made by allowing additional information to be used as retrieval cues. In the current framework, the goal of the task (i.e. choosing the option with the higher criterion value) is used as the retrieval cue. However, as noted in the introduction, other information could also serve as retrieval cues. For example, the actual options being compared could prompt the retrieval of different cues, which might happen if certain options are more strongly associated with particular cues. The cues themselves could also be used as retrieval cues. This would likely occur if cues are intercorrelated, such that the value of one cue provided information about the value of another cue. In this case, retrieving one cue might prompt the retrieval of a related cue.

Moving toward decision specific predictions could address another potential shortcoming of the accessibility framework. As noted in the discussion of single strategy models above, one of their strengths is in accounting for evidence that people are sensitive to the consistency of cue information, meaning when all cues do not support the same option (Dummel et al., 2016; Glöckner et al., 2010; Glöckner & Hodges, 2011). For example, studies found that people continued searching cues after finding inconsistent information (Dummel et al., 2016; Glöckner & Betsch, 2012). PCS and EAM both make predictions about the effects of inconsistent information. PCS assumes that when less valid cues are inconsistent with more valid cues confidence will be lower and decision times will increase because the consistency maximization process is more difficult. Similarly, EAM assumes that cue search will be more extensive if less valid cues are inconsistent with more valid cues because it will take longer to reach the threshold. Although consistency affects the confidence within the accessibility framework, it does not affect the cue generation process directly for the reasons listed above. None of the experiments in this dissertation were directly testing effects of information consistency, but it is possible that information consistency was affecting participant's decisions. Thus, allowing the accessibility framework to adjust search based on the specific decision being made, may improve its ability to account for participant performance in a psychologically plausible way.

By making these changes to the accessibility framework, it should be able to address criticisms of the current frameworks and solve the shortcomings uncovered in this dissertation. Unlike the current frameworks, the accessibility model is not validity-based and is able to account for the effects of memory on cue-based inferences. The accessibility framework resolves the criticisms of the adaptive toolbox by eliminating the strategy selection problem, providing a falsifiable model, and being able to account for individual differences. It also addresses the criticisms of the single strategy models by fully specifying search and not being overparameterized. The changes discussed above will improve its ability to account for the amount of information people use when making decisions and allow it to adapt to specific decisions. Because the framework assumes cue use is dependent on the accessibility of the cues in memory, the accessibility framework is more plausible and parsimonious than the previous accounts. Moreover, it provides a foundation for future research looking more closely at the role of memory in cue-based inferences, both in experimental and applied settings.

As discussed above, much of the work within cue-based inferences are based on the seminal work of Payne et al. (1988) who studied how people adapt to different decision environments. This dissertation also finds evidence for adaptive decision-making but suggests that the adaptations may be more nuanced than changing search or integration strategies. Understanding that people tend to use most cues available them, even when additional cues are not necessary to arrive at the correct decision, can help researchers understand why certain decision errors are made and can help researchers work toward improving these errors. For example, a decision aid could be designed to only show information known to be important for the decision at hand and not include additional information simply because the decision-maker requests it.

CHAPTER 7: CONCLUSION

Although the studies presented in this dissertation do not conclusively support the accessibility framework for cue-based inferences, they still provide several important contributions to the literature. First, they extend previous research showing that memory affects cue-based inferences, challenging frameworks that are based on validity only. Second, they extend research on adaptive decision-making by showing that people are sensitive to the decision environment but that this does not always result in changes to both decision outcomes and decision processes. Third, many of the results directly challenge assumptions of the adaptive toolbox, and Take-the-Best in particular, because participants did not often appear to use noncompensatory strategies. Overall, the accessibility framework provides a promising foundation for explaining how people make cue-based inferences, but further research is necessary to better understand how people search for cues, particularly how they decide to stop searching.

REFERENCES

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(1), 107–129. https://doi.org/10.1037/0278-7393.33.1.107

Betsch, T., & Glöckner, A. (2010). Intuition in judgment and decision making: Extensive thinking without effort. *Psychological Inquiry*, *21*(4), 279–294. https://doi.org/10.1080/1047840X.2010.517737

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, *6*(2), 173-189.

Bröder, A. (2000). Assessing the empirical validity of the "Take-The-Best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1332–1346. https://doi.org/10.1037//0278-7393.26.5.1332

Bröder, A. (2003). Decision making with the "adaptive toolbox": influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *29*(4), 611–625. https://doi.org/10.1037/0278-7393.29.4.611

Bröder, A., & Gaissmaier, W. (2007). Sequential processing of cues in memory-based multiattribute decisions. *Psychonomic Bulletin & Review*, *14*(5), 895–900. https://doi.org/10.3758/BF03194118

Bröder, A., & Schiffer, S. (2003). Take the best versus simultaneous feature matching: probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology. General*, *132*(2), 277–293. https://doi.org/10.1037/0096-3445.132.2.277

Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *32*(4), 904–918. https://doi.org/10.1037/0278-7393.32.4.904

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459. https://doi.org/10.1037/0033-295X.100.3.432

Dieckmann, A., & Rieskamp, J. (2007). The influence of information redundancy on probabilistic inferences. *Memory & Cognition*, *35*(7), 1801–1813. https://doi.org/10.3758/BF03193511

Dimov, C. M. (2017). How to implement HyGene into ACT-R[*]. *Journal of Cognitive Psychology*, *0*(0), 1–14. https://doi.org/10.1080/20445911.2017.1394863

Dougherty, M. R., Franco-Watkins, A. M., & Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychological Review*, *115*(1), 199–213. https://doi.org/10.1037/0033-295X.115.1.199

Dougherty, M. R., Thomas, R., & Franco-Watkins, A. M. (2008). Postscript: Vague

Heuristics Revisited. *Psychological Review*, *115*(1), 211–213. https://doi.org/10.1037/0033-295X.115.1.211

Dummel, S., Rummel, J., & Voss, A. (2016). Additional information is not ignored: New evidence for information integration and inhibition in take-the-best decisions. *Acta Psychologica*, *163*, 167–184. https://doi.org/10.1016/j.actpsy.2015.12.001

Gigerenzer, G., & Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science*, *1*(1), 107–143. https://doi.org/10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way : Models of bounded rationality. *Psychological Review, 103*(4), 650–669. https://doi.org/10.1093/acprof

Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011). *Heuristics: The foundations of adaptive behavior*. New York: Oxford University Press.

Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple Heuristics That Make Us Smart*. *New York: Oxford University Press*.

Glöckner, A., & Betsch, T. (2008a). Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgment and Decision Making*, *3*(3), 215–228. https://doi.org/10.2139/ssrn.1090866

Glöckner, A., & Betsch, T. (2008b). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1055. https://doi.org/10.1037/0278-7393.34.5.1055

Glöckner, A., & Betsch, T. (2012). Decisions beyond boundaries: When more information is processed faster than less. *Acta Psychologica*, *139*(3), 532–542. https://doi.org/10.1016/j.actpsy.2012.01.009

Glöckner, A., Betsch, T., & Schindler, N. (2010). Coherence Shifts in Probabilistic Inference Tasks. *The Journal of Behavioral Decision Making*, *23*, 439–462. https://doi.org/10.1002/bdm.668

Glöckner, A., & Bröder, A. (2011). Processing of recognition information and additional cues: A model-based analysis of choice, confidence, and response time. *Judgment and Decision Making*, *6*(1), 23–42.

Glöckner, A., Hilbig, B. E., & Jekel, M. (2014). What is adaptive about adaptive decision making? A parallel constraint satisfaction account. *Cognition*, *133*(3), 641–666. https://doi.org/10.1016/j.cognition.2014.08.017

Glöckner, A., & Hodges, S. D. (2011). Parallel constraint satisfaction in memory-based decisions. *Experimental Psychology*, *58*(3), 180–195. https://doi.org/10.1027/1618-3169/a000084

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological review*, *93*(3), 258-268.

Hausmann, D., & Lage, D. (2008). Sequential evidence accumulation in decision making: The individual desired level of confidence can explain the extent of information acquisition. *Judgment and Decision Making Journal*, *3*(3), 229–243.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551. https://doi.org/10.1037/0033-295X.95.4.528

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*(1), 1-55.

Lawrence, A., Thomas, R. P., & Dougherty, M. R. (2018a). Integrating fast-and-frugal heuristics with a model of memory-based cue generation. *Journal of Behavioral Decision Making*.

Lawrence, A., Thomas, R. P., & Dougherty, M. R. (2018b). Long-term serial position effects in cue-based inferences. *Plos One, 13*(7).

Lange, N. D., Thomas, R. P., Buttaccio, D. R., Illingworth, D. A., & Davelaar, E. J. (2013). Working memory dynamics bias the generation of beliefs: The influence of data presentation rate on hypothesis generation. *Psychonomic Bulletin & Review*, *20*(1), 171-176.

Lange, N. D., Thomas, R. P., & Davelaar, E. J. (2012). Temporal dynamics of hypothesis generation: the influences of data serial order, data consistency, and elicitation timing. *Frontiers in Psychology*, *3*, 215.

Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the “take the best” and the “rational” models. *Psychonomic Bulletin & Review*, *11*(2), 343–352. https://doi.org/10.3758/BF03196581

Lee, M. D., & Zhang, S. (2012). Evaluating the coherence of Take-the-best in structured environments Take-the-best. *Judgment and Decision Making*, *7*(4), 360–372.

Marewski, J. N. (2010). On the Theoretical Precision and Strategy Selection Problem of a Single-Strategy Approach: Comment on Glockner, Betsch, and Schindler (2010). *The Journal of Behavioral Decision Making*, *23*, 463–467. https://doi.org/10.1002/bdm

Marewski, J. N., & Link, D. (2014). Strategy selection: An introduction to the modeling challenge. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(1), 39–59. https://doi.org/10.1002/wcs.1265

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: an ecological model of strategy selection. *Psychological Review*, *118*(3), 393–437. https://doi.org/10.1037/a0024143

Newell, B. R. (2005). Re-visions of rationality? *Trends in Cognitive Sciences*, *9*(1), 11–15. https://doi.org/10.1016/j.tics.2004.11.005

Newell, B. R., & Bröder, A. (2008). Cognitive processes, models and metaphors in decision research. *Judgment and Decision Making*, *3*(3), 195–204.

Newell, B. R., & Lee, M. D. (2011). The right tool for the job? Comparing an evidence accumulation and a naive strategy selection model of decision making. *Journal of Behavioral Decision Making*, *24*, 456–481. https://doi.org/10.1002/bdm

Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision making. *Journal of Behavioral Decision Making*, *17*, 117–137. https://doi.org/10.1002/bdm.465

Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision making. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *29*(1), 53–65. https://doi.org/10.1037/0278-7393.29.1.53

Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone "takes-the-best." *Organizational Behavior and Human Decision Processes*, *91*(1), 82–96. https://doi.org/10.1016/S0749-5978(02)00525-3

Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, *29*(4), 559–586. https://doi.org/10.1207/s15516709cog0000_14

Payne, J., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*(3), 522–534. https://doi.org/10.1037//0278-7393.14.3.534

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: University of Cambridge.

Payne, J. W., Bettman, J. R., & Luce, M. F. (1996). When Time Is Money: Decision Behavior under Opportunity-Cost Time Pressure. *Organizational Behavior and Human Decision Processes*, *66*(2), 131–152. https://doi.org/10.1006/obhd.1996.0044

Peterson, C. R., & DuCharme, W. M. (1967). A primacy effect in subjective probability revision. *Journal of Experimental Psychology*, *73*(1), 61.

Platzer, C., & Bröder, A. (2012). Most people do not ignore salient invalid cues in memory-based decisions. *Psychonomic Bulletin & Review*, *19*(4), 654–661. https://doi.org/10.3758/s13423-012-0248-4

Platzer, C., Bröder, A., & Heck, D. W. (2014). Deciding with the eye: How the visually manipulated accessibility of information in memory influences decision behavior. *Memory & Cognition*, *42*(4), 595–608. https://doi.org/10.3758/s13421-013-0380-z

Rakow, T., Newell, B., Fayers, K., & Hersby, M. (2005). Evaluating three criteria for establishing cue-search hierarchies in inferential judgment. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(5), 1088–1104.

https://doi.org/10.1037/0278-7393.31.5.1088

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, *127*(2), 258–276. https://doi.org/10.1016/j.actpsy.2007.05.004

Rieskamp, J., & Otto, P. E. (2006). SSL: A Theory of How People Learn to Select Strategies. *Heuristics: The Foundations of Adaptive Behavior*, *135*(2), 207–236. https://doi.org/10.1093/acprof:oso/9780199744282.003.0011

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*(4), 734–760. https://doi.org/10.1037/0033-295X.103.4.734

Scheibehenne, B., Rieskamp, J., & Wagenmakers, E.-J. (2013). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological Review*, *120*(1), 39–64. https://doi.org/10.1037/a0030777

Söllner, A., & Bröder, A. (2015). Toolbox or adjustable spanner? A critical comparison of two metaphors for adaptive decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 215–237. https://doi.org/10.1037/xlm0000162

Söllner, A., Bröder, A., Glöckner, A., & Betsch, T. (2014). Single-process versus multiple-strategy models of decision making: Evidence from an information intrusion paradigm. *Acta Psychologica*, *146*(1), 84–96. https://doi.org/10.1016/j.actpsy.2013.12.007

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review, 115*(1), 155–

185. https://doi.org/10.1037/0033-295X.115.1.155

Wixted, J. T. (2004). The Psychology and Neuroscience of Forgetting. *Annual Review of Psychology*, *55*(1), 235–269. https://doi.org/10.1146/annurev.psych.55.090902.141555

Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science*, *18*(2), 133–134. https://doi.org/10.1111/j.1467-9280.2007.01862.x

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, *2*(6), 409–415.