STATISTICAL MANIPULATION AND CONTROL STRATEGIES OF THE N-BACK TASK

By

Jason Ralph

A Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: COGNITIVE SCIENCE

Approved by the
Examining Committee:

Wayne Gray, Thesis Adviser

Michael Shoelles, Member

Brett Fajen, Member

Michael Kalsher, Member

Christopher Chabris, Member

Rensselaer Polytechnic Institute
Troy, New York

November 2014
(For Graduation December 2014)

Contents

List of Tables

List of Figures

ACKNOWLEDGMENT

I'd like to acknowledge all of the people who have made this project possible. First, the comments and help from my advisor, Wayne Gray. Because I began the program at an older age than most graduate students, Dr. Gray's patience and understanding regarding personal matters was absolutely critical and made the completion of this dissertation possible. Also, our conversations over the years provided immeasurable inspiration that have shaped the direction and value of this dissertation.

Second, the contributions of my committee cannot be overstated. Michael Shoelles deserves special mention. There is no way this project could have been completed without Dr. Schoelles' expertise in writing the experiment code for all three experiments. I credit Brett Fajen with stimulating important thinking regarding the generalization of theoretical constructs beyond the n-back. Mike Kalsher has been an invaluable resource in developing and applying statistical tests that are not commonly used in the n-back literature. Finally, I d́ like to thank Chris Chabris for serving on the committee and providing the perspective of a researcher who has used the n-back in the past.

I would also like to thank the members (past and present) of the cogworks lab. Discussions over the years with Mike Brudzinski, John Lindstadt, Ryan Hope, and Kevin Berry helped to stimulate my thinking and focus the research questions considered in this work. In addition, this dissertation could also not have been completed without the amazing dedication of my "army of undergraduates," Anisha Smith, Charles Roohan, Luna King, and Stefanie Turk, who helped to administer many of the experimental sessions. I wish them the best in their future endeavors.

Finally, I would like to thank two very special people in my life. There is no possible way this thesis could have been completed without the inspiration, patience, support and love of my wonderful wife Lauren and daughter Teagan. I very much appreciate all the sacrifices they have made over the past 7 years and look forward to celebrating life going forward!!

ABSTRACT

The impact of task environments on cognitive strategies are well known and have formed the basis of many important findings in cognitive science. However, these effects are often overlooked in research design. The n-back, thought to measure working memory and executive control capacity, is especially prone to misinterpretation because of a lack of standardization in how researchers administer the task. This dissertation chronicles a set of experiments investigating experimental design variation along several dimensions. The results showed that differences in stimulus frequency distributions and the ratio of targets to lures cause variation in performance by stimulating strategic differences. Variations in n-back performance across different populations or different n-back paradigms are often interpreted in terms of resource capacity or strategic views of memory, attention, and cognitive control. Hence, the results of this work highlight the importance of understanding the connection between strategy and the environment when applying cognitive theory to tasks like the n-back.

Chapter 1
Background

1.1 Introduction

In addition to managing top-down control of our actions, our executive control functions make it possible to adjust quickly to small changes in our environment. Drivers slow down and leave more room between cars in rainy conditions. Hikers more carefully place each step on rocky terrain. Experienced skiers adjust their lines to match the iciness of the slope. The increased likelihood of errors in these environments lead to changes in the use of cognitive, perceptual, and motor processes, oftentimes without conscious awareness.

Herbert Simon described an ant traveling across a sand dune as an example of how the environment affects cognition (Simon, 1969). Just as we cannot predict the movements of Simon's ant without considering the topography of the sand dune, we cannot understand the cognitive processes that underlie observable performance in more complex tasks without considering the intricacies of the task environment. The environment plays no less of a role in affecting complex human cognition. Our ability to use top-down control to guide our actions sometimes makes us oblivious to the importance of environmental factors.

Researchers must consider the influence of experimental design on the cognitive strategies used for task performance. Small changes in the task environment can induce major adjustments in the processes used to perform a task, and elicit great differences in performance. Without understanding these interactions even common paradigms can lead to scientific confusion rather than provide a source to isolate and study cognitive processes.

The current work focuses on specific statistical features of the task environment that can change the way people approach a task. In general, people perform tasks using either a proactive or reactive control strategy. Proactive control involves the activation of task context in preparation for upcoming stimuli, and reactive control refers to the transient activation of task context in response to stimuli. The Dual

Mechanisms of Control (DMC) theory (Braver, Gray, & Burgess, 2007; Braver, 2012) argues that factors in the environment such as the likelihood of errors and the amount of required information processing determine the control mode used. Changes in control strategy affect many of the resources the brain uses to perform tasks (e.g. memory, and attention), so it is critical that we understand how factors in the environment, such as the probability of various events occurring, affect the mode of control used.

For instance, I often eat lunch at a "make your own" burrito restaurant, where you choose various toppings and an employee prepares your order on the spot. One day I could not decide, so I just ordered the same thing as the customer in front of me. The employee looked down at the containers of toppings, and then admitted in a bewildered voice that he had no idea what he had just prepared. Although I assume his memory capacity is large enough to have remembered the list of 4 or 5 items he had just put on the previous customer's burrito, for some reason that information was no longer available just seconds later.

We can explain this phenomena by analyzing the task environment. First, there are relatively few toppings, so we can assume the employee uses each of these often. It is likely that proactive interference is quite high as every combination of toppings would be equally familiar. Second, there is no reason to believe that people tend to order the exact same combination as the person directly in front of them, so there is no use to remember an order after it is made. The statistical nature of the environment played a role in the ability to remember a relatively small number of items over a very short time. It is likely that had I alerted him before he helped the previous customer that I would be ordering the same thing, he would have used proactive control and been able to remember the order.

Subjects in working memory experiments are generally biased toward a proactive control mode. Usually we expect them to remember items for short periods of time, the items are useful, and no environmental cues predict the correct response. This bias toward proactive control does not always exist in the real world. Our forgetful employee was not in an experimental lab. He was acting in a real world environment that biased him toward a reactive strategy.

We are often unaware of the external factors that affect our behavior. Reder and colleagues found that people react to conditional probabilities in a spatial localization task which involves making quick responses to targets presented in the presence or absence of a distractor (Reder, Weber, Shang, & Vanyukov, 2003). Distractors appeared in one location 60% of the time, 10% of the time in each of three other positions, and was absent 10% of the time. People learned to suppress responses when the target appeared in a location that usually contained a distractor. Interestingly, despite clear differences in behavior, subjects generally reported no awareness of a difference in the likelihood of an item appearing in a given location.

Crowley, Shrager and Siegler discovered a similar phenomenon, noting that strategies often are developed and employed without conscious awareness (Crowley, Shrager, & Siegler, 1997). In their study, children learned to solve math problems of the form a + b - a by canceling out the a and -a and responding with b. An analysis of response times showed that this strategy was actually used several trials before the children reported its use. Thus subconscious strategy selection occurs in domains more complex than simple reactions to stimuli, as was the case in the Reder study.

1.1.1 The *n*-back task

The current work focuses on the *n*-back sequential memory task in order to better understand the interactions between variations in a task environment and strategy development. The history of the *n*-back provides an ideal example of the confusion that can be caused by small changes in a task environment. Researchers have developed no standards for how the task is to be administered: the stimuli used, time intervals between presentation, and the statistical properties of stimulus sequences vary wildly between experiments, as the design variations between *n*-back studies illustrates (and will be discussed in detail in chapter 2) (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Brouwer et al., 2012; Braver, Cohen, et al., 1997; Owen, McMillan, Laird, & Bullmore, 2005).

Although researchers intentionally manipulate the task environment in order to test hypotheses regarding various cognitive processes, many studies also feature unintentional variation of the environment. Unintentional variation should cause concern, especially for memory intensive tasks that are particularly susceptible to small changes in the task environment. Variation in the statistical properties of the n-back is particularly important because the relative likelihood of a given event occurring can impact both memory retrieval and interference control (Anderson, Bothell, et al., 2004; Botvinick, Braver, Barch, Carter, & Cohen, 2001). These are the two cognitive functions thought to be most responsible for variation in performance on the n-back task (Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011; Burgess, Gray, Conway, & Braver, 2011).

Despite inconsistencies in experimental procedures, the n-back has been used extensively as a measure of working memory capacity, especially in fMRI research (see Owen et al., 2005 for a review). Many of these studies cite brain activity recorded during the n-back as reflecting neural correlates of working memory. As the experiments in this work demonstrate, strategic variation caused by experimental design differences render these conclusions dubious at best.

The n-back is a sequential memory task in which people view a series of items (e.g. letters) one at a time and determine if each is the same or different as the letter presented a specific number of items in the past. For clarity and standardization, all of the examples and discussion herein will focus on a 3-back version. Thus the target item always matches the item presented three trials in the past. Consider the presentation sequence (B A C **B** D **C** *D*). The bolded **B** and **C** are matches to the third previous letter. Care must be taken to avoid false alarms on trials that match the 2nd or 4th previous letters, commonly referred to as recent lure trials. For instance, the last letter (*D*) matches the 2nd previous but not the critical 3-back item.

Accurate performance requires people to correctly retrieve the 3rd previous letter while avoiding false alarms on recent lure trials. The statistical properties of the task strongly affect both of these abilities. The ease of retrieving an item from memory relates to the relative recency and frequency of its presentation (Anderson & Lebiere, 1998; Anderson, Bothell, et al., 2004). Items seen frequently will be easier to remember than those rarely presented. People may become primed to recognize high frequency items as matches regardless of how many items back they were last presented.

Differences in the presentation frequency of individual stimuli can bias people toward an availability heuristic much like the one discussed by Gigerenzer and Brighton (2009). When asked to determine which of two foreign cities has a higher population, people tend to rely on the availability of the two cities in their memory (e.g. Berlin is more familiar to me so it must have more people than Stuttgart). Simply put, the availability of information in memory becomes the deciding factor. The activation of items in memory can also be helpful in the n-back task, dependent upon the statistics of the presentation sequence. A reactive strategy based on activation works best when frequently seen items tend to be matches. On the other hand, sequences with a high number of lure trials make activation less useful.

Correct rejection of lures requires effective executive control. In order to avoid errors people must be able to recognize a conflict between the high familiarity of the current item and the controlled recollection of the target (3-back) item. The likelihood of a difficult (conflicting) stimulus affects the ability to recognize and resolve this conflict (Botvinick et al., 2001; Yeung, Botvinick, & Cohen, 2004). In traditional conflict paradigms such as the Stroop and flanker tasks, incongruent trials cause errors and increased response times. Incongruent trails are those in which the stimulus simultaneously primes two competing responses. Conversely, congruent trials prime only one response. The difference in performance between congruent and incongruent trials becomes less pronounced as the ratio of congruent to incongruent trials approaches 1. This change in the impact of conflict can be explained by a subtle change in control mode from reactive to proactive. As conflict becomes more likely, people begin to prepare for it ahead of time.

Lure trials in the n-back simultaneously prime "match" and "no-match" responses because of their high activation. Applying the findings from the conflict monitoring literature, we can expect the likelihood of lure trials to affect the use of proactive and reactive control in the n-back. If we present people with n-back sequences which contain numerous recent lure trials, we expect them to prepare ahead of time by actively maintaining the target item. By contrast, we expect that nback sequences with rare lure trials will bias people toward an availability based reactive strategy.

Interestingly, the n-back creates a type of conflict not found in other control paradigms. Conflict in the Stroop and flanker tasks occurs completely within one trial, as various parts of the stimulus activate competing responses. In the Stroop task, for instance, the word GREEN written in a red font color creates conflict by simultaneously activating Red and Green responses. In the flanker task, a stimulus such as <<><< primes both left and right responses. By contrast, n-back conflict is caused by between-trial factors. A lure trial is difficult not because of the letter itself, but because that letter also appeared two trials in the past. Thus the n-back is a unique task (with respect to cognitive control) because between-trial sequence factors, often overlooked in the design of experiments, determine the degree and type of cognitive control required for accurate task performance.

In addition to between-trial factors, the strategy used to make comparisons to past items can lead to differences in control mode. In order to maximize accuracy on match trials, many models of the n-back use a rehearsal strategy that increases the activation of the last n items in the sequence (Harbison, Atkins, & Dougherty, 2011; Juvina & Taatgen, 2007). Rehearsal enhances the activation of the target item, making an accurate response more likely. Unfortunately, excessive rehearsal also increases the activation of items surrounding the target, making lure trials even more difficult to reject. It is easy to see a vicious cycle in which rehearsal leads to lure interference which must be overcome by performing more rehearsal and so on. Szmelac found this effect in his 2011 study in which longer inter-stimulus intervals led to increased response times for lure trials (Szmalec et al., 2011). Thus more rehearsal led to more conflict.

Because of increased conflict that can result from rehearsal, variations in the strategy used to memorize past items and integrate new ones are likely to cause variation in n-back task performance. Although the n-back literature does not tend to focus on strategies per se, several have been identified (Chatham et al., 2011; Juvina & Taatgen, 2007; Harbison et al., 2011). These studies do propose strategies and provide models which explain some of the variance in the data, but they do not consider the existence of other potential strategies, nor do they investigate variation within a given strategy.

This work will reveal the n-back to be a surprisingly rich task that integrates many issues of great importance to the cognitive science community. I propose that the memory strategy used, the control mode engaged, and the statistical probabilities inherent in the task environment will interact with each other to yield differences in performance that cannot be explained without considering all three factors.

In the following section I will provide a brief review of current understanding of the memory and control issues involved in the n-back. In chapter 2, I will investigate the variation in sequences found in the n-back literature and discuss how these factors might affect strategy selection and use in the n-back. In the remaining chapters I will report the results of a series of experiments testing the effect sequence factors have on n-back strategies and resulting performance.

1.2 A Brief Review of Working Memory

The n-back is used by some to measure working memory capacity, however several studies have cast doubt on its validity as a WM measure (Kane, Conway, Miura, & Colflesh, 2007; Redick et al., 2012). In fact, researchers have debated the nature and capacity of working memory for decades. In this section I will provide a brief overview of the relevant theories, focusing specifically on what they would predict in relation to the n-back.

In describing his famous model of working memory, Alan Baddeley introduced the concept of the phonological loop, which holds and rehearses verbal information. (Baddeley & Scott, 1971; Baddeley, 1992b; Baddeley, 1992a, 1984).

> The phonoloical loop is assumed to have two components, a brief speech- based store that holds a memory trace that fades within approximately 2 secs, coupled with an articulatory control process. This process, which resembles sub-vocal rehearsal, is capable of maintaining the material in the phonological store by a recycling process, and in addition is able to feed information into the store by a process of subvocalization. (Baddeley, Emslie, Kolodny, & Duncan, 1998, p.284)

The concepts of decay and rehearsal are central to Baddeley's theory and those that followed (Miyake & Shah, 1999). Verbal information is thought to decay quickly, but rehearsal can refresh the activation, much like a juggler tossing items in the air (Nairne, 2002). The juggler metaphor is apt in many regards. As Nairne points out, "The juggler is able to maintain a set of activated items to the extent that each can be caught and re-tossed before gravity reduces it to an irretrievable state." (Nairne, 2002, p. 55). The capacity of verbal working memory results from a trade-off between decay and the efficiency of the rehearsal system that keeps items active. Despite the reputation of Baddeley's model, experimental results have provided data that are not easily accounted for by a simple decay/rehearsal account of working memory (see Nairne, 2002 for a review). Although decay and rehearsal may play a role in the ability to maintain memory, other factors such as interference from intervening stimuli may be more likely to account for experimental data.

Many researchers conceptualize working memory as a multi-item store, but McElree and colleagues argued that the mind only actively maintains the current focus of attention, and retrieves all other items from long term memory when needed (McElree, 2001). Neural evidence backs up McElree's claims as the activity of the hippo-campus during retrieval of items in working memory replicates activity recorded when people remember long-term memories (Oztekin, Davachi, & McElree, 2010). Only the most recent item elicits the type of neural activity generally associated with working memory function.

Anderson's rational analysis of memory (incorporated into the ACT-R cognitive architecture) complements McElree's view. There is no distinction in ACT-R between working, short, and long-term memory. The ability to retrieve an item from memory is determined simply by a time based decaying activation function, reflecting the relative recency and frequency of use. ACT-R is more likely to successfully retrieve items that have been seen recently and often due to increased activation values (Anderson, Bothell, et al., 2004).

In order to hold items in an easily retrievable state (such as the last 3 items presented in the n-back task) the cognitive system must periodically rehearse the needed items to keep their activation from decaying below a retrieval threshold.

Thus performance in the n-back task may reflect the control necessary to perform these rehearsals in the absence of an explicit cue to do so. For instance, models of the Stroop task often use goal rehearsals to "remind" oneself to focus on color in order to override the natural tendency to respond to the word itself (Botvinick et al., 2001; Bailey, West, & Anderson, 2010; Juvina & Taatgen, 2007).

Juvina and Taatgen (2007) had subjects perform both the n-back and stroop tasks and found that Stroop performance was better for subjects who used rehearsal in the n-back than for those who primarily used familiarity judgments. Despite clear differences between these tasks, the use of proactive control (rehearsing the target item in the nback and rehearsing the correct goal in the Stroop task) is a common component that can explain much of the variation in performance of these tasks.

1.3 Dual Mechanisms of Control

Prominent memory theorists such as Baddeley, McElree, and others focus on the processes the brain uses to hold information active, but ignore questions of how and when the brain chooses to use these processes. The Dual Mechanism of Control (DMC) theory argues that the activation of specific information ahead of time, known as proactive control, is used sparingly, on an "as needed" basis. Neural evidence suggests that representations are held in pre-frontal cortex (PFC) and kept active in anticipation of stimulus presentation in a variety of tasks (Braver, Gray, & Burgess, 2007; Braver, 2012). When not in a proactive state, memory representations are recruited only after stimulus presentation. This is known as reactive control, in which attentive mechanisms are initiated when a stimulus activates multiple competing responses.

It is important to note that the term proactive does not simply refer to any consciously chosen strategy. For the purpose of this work (and the DMC theory generally), proactive control refers specifically to the activation of task context in preparation for upcoming stimuli. Consciously chosen strategies only qualify as proactive control if they activate task context in preparation of an upcoming stimulus.

Some consciously chosen strategies can have inherently reactive instead of proactive control qualities. For instance, many models of the n-back make use of sub-vocal rehearsal. Rehearsal increases the activation of the past three items and leads many to consider this rehearsal to qualify as a proactive strategy.

However, rehearsal only fits the proactive definition if it involves actively holding the 3rd previous item "in mind" in preparation for each successive item. If the rehearsal simply increases the activation of multiple items and the subject retrieves the most active item from memory in response to a new stimulus, then this rehearsal fits the definition of a reactive strategy.

Furthermore, people often engage proactive control without conscious awareness. Many examples of this occur in the literature including findings from the AX-CPT (Braver, Gray, & Burgess, 2007) and Sternberg memory task (Speer, Jacoby, & Braver, 2003). These studies found the use of proactive control often depends upon sub-conscious analysis of the task environment.

The relative use of proactive and reactive control is subject to individual differences and is strongly correlated with differences in task performance. For instance, young adults are more likely to use proactive control while older people and schizophrenic populations tend to be reactive (Braver, Gray, & Burgess, 2007; Braver, 2012). In addition to individual differences, control modes are sensitive to various factors of the task environment including expected memory load, time constraints, the predictive value of cues, and the likelihood of interference. The following sections will examine evidence for each of these effects.

1.3.1 Expected Memory Load

The use of proactive control tends to decrease as the number of items to remember and the resulting cognitive load increases. It is somewhat counter-intuitive that people would use less proactive control as load increases. Perhaps the learned effectiveness of a strategy may help determine if it is chosen for use. People may learn that there is an upper limit to the number of items they can successfully rehearse, and use reactive control when a task exceeds that limit.

Speer and Braver presented subjects with a version of the Sternberg memory recognition task in which subjects viewed lists of words and then determined if a specific word was present in the most recent set. One group of subjects viewed sets of 8 words, and another viewed sets with 3 words. Researchers presented subjects in both groups sets with 6 words on 20% of trials. Analysis of fMRI scans showed that subjects used proactive control when 3-item lists were common, but used reactive control when they expected 8 items. Subjects in both conditions retained their dominant control mode when presented with the 6-word sets, suggesting that people use control modes that are matched to learned expectations of memory load (Braver, Gray, & Burgess, 2007; Speer et al., 2003).

As discussed earlier, Oztekin and McElree argued against proactive working memory process on the basis of fMRI data that shows no sign of active maintenance of previously experienced stimuli. However, McElree's study (Oztekin et al., 2010) used relatively long study lists (6 words). Given the evidence presented by Speer (2003), Oztekin may have biased subjects toward reactive control because longer study lists result in higher expected load. Thus the lack of evidence of proactive processes in their study may say more about the effect of task parameters such as list length on the control mechanisms used in the task than about the nature of working memory itself.

1.3.2 Time Constraints

The time between stimulus presentations can also serve to bias control modes. Szmalec (2011) manipulated inter-stimulus intervals (ISI) to demonstrate this effect. Short ISIs (2000 ms) allow less time for rehearsals than long ISIs (4000 ms). Szmelac found evidence of increased activity of items in the long ISI condition, resulting in more conflict as proactive rehearsals led to difficulty in correctly rejecting lure trials. Response times were longer for lure trials during long ISI blocks than during short ISI blocks. On the other hand, target accuracy increased slightly, so proactive rehearsals in the n-back has costs (increased lure interference) and benefits (increased target accuracy).

Szmelac's finding of an effect of ISI on control strategy is interesting in light of McElree's 2001 study of sequential working memory (McElree, 2001). McElree used a novel procedure in which 6-13 items were presented in sequence. After a random number of items (between 6 and 13), an auditory probe prompted subjects to respond as to whether the most recent item matched the nth previous letter. McElree reported evidence that subjects performed memory search operations upon viewing the probe stimulus, as opposed to making a simple comparison to the nth previous item held in focal attention.

McElree concluded that people perform the *n*-back reactively, searching their memory after presentation of the probe stimulus. However, this conclusion must be tempered by the fact that McElree presented the list of stimuli very quickly, allowing only 900 ms between stimuli in the 3-back condition (1 and 2 back stimuli were presented even faster). Given Szmalec's results, it is not surprising that McElree found evidence for a reactive process as there was simply not enough time to proactively update and rehearse items between stimulus presentations. Again, the task environment played a major role in control strategy choice.

1.3.3 Cue Predictability

Cue predictability is the most commonly researched factor that affects control. Research using the AX version of the continuous performance task (AX-CPT) has found consistent effects of cue predictability. As in the *n*-back, subjects view a series of letters presented at fixed time intervals. The experiment presents items in pairs, choosing a letter from the first half of the alphabet as a cue, and following an ISI (usually 3000-5000ms), displays a probe chosen from the 2nd half of the alphabet. Subjects must respond yes only if the current letter (probe) is an X and the previous letter (cue) was an A. Many researchers present these AX trials 70% of the time, creating a situation where an A cue strongly predicts an upcoming X probe (Braver, Gray, & Burgess, 2007; Braver, Paxton, Locke, & Barch, 2009). Healthy college aged subjects tend to engage proactive control upon viewing an A, evidenced by sustained activation of the lateral prefrontal cortex over the ISI between the cue and probe (Braver, Gray, & Burgess, 2007; Braver, 2012). Elderly and Schizophrenic populations do not tend to exhibit this type of brain activity, suggesting that they rely on reactive control.

In addition to differences in brain activity, task performance differs between groups in systematic ways. The proactive (younger) group tends to be faster and more accurate on the common AX trials, but make more errors on rare AY trials when compared to the reactive (older and schizophrenic) group. The behavioral results in these studies suggest a cost/benefit trade off in control mode selection. Proactive control in the AX-CPT leads to increased accuracy on the predicted X trials following A cues at the expense of an increased error rate on trials where another letter follows an A. All else being equal the cost/benefit analysis breaks in favor of proactive control as there are 700% more AX than AY trials. Interestingly, the control mode can be altered by instructing subjects to focus on AY accuracy instead of speed, resulting in reactive control as evidenced by fMRI and behavioral changes (Braver, Paxton, et al., 2009).

It is not exactly clear how cue predictability may work in the n-back. Each stimulus is a cue, with the probe trial appearing n trials in the future. If all possible stimuli are chosen an equal number of times, none of the stimuli would trigger proactive control due to cue predictability because there is no correlation between the cue and its appearance as a probe n trials in the future. On the other hand, sequences with a skewed frequency distribution present a specific stimuli often on match trials. Consider a 3-back task with the following sequence: (A B A A D A A A C B G D). After the initial 3 letters (A B A), 3 of the next 4 A's are matches. By contrast, none of the other 4 letters (D C B G) were chosen as matches. A reactive strategy could be very successful here, as the frequently seen letters would become more available in memory and easier to recognize as matches.

1.3.4 Likelihood of Interference

As discussed previously, interference control plays a large role in the ability to correctly reject lure trials in the n-back (Burgess & Braver, 2010; Burgess, Gray, et al., 2011; Szmalec et al., 2011). Effective reactive control is necessary to avoid committing false alarm errors because lure trials simultaneously activate match and no-match responses. This occurs because n-back performance relies both on judgments of familiarity and serial position (Harbison et al., 2011; Juvina & Taatgen, 2007; Szmalec et al., 2011). These judgments lead to conflicting responses on lure trials.

In other control paradigms, such as the flanker and Stroop tasks, proactive control generally reduces the amount of conflict caused by incongruent stimuli. In the flanker task the target center arrow is flanked by either congruent or incongruent arrows. In the Stroop task, color words (e.g. red or green) are displayed in either congruent or incongruent font colors. Each of these tasks induce conflict between competing responses. Proactive control in the flanker task results in increased attentional focus on the central arrow, reducing the impact of the flanker arrows. Proactive control in the Stroop task reduces the likelihood of responding to the word presented, resulting in higher accuracy when attempting to name the font color. The control mechanism used in these tasks is strongly dependent on the proportion of incongruent trials. Conditions with rare incongruent trials lead to reactive control, and conditions with common incongruent trials prime proactive control (Botvinick et al., 2001; Bailey et al., 2010).

Burgess et al (2010) found that control in working memory tasks follows the same pattern. Subjects performed a recent negatives task which presented 5 item memory lists, followed by a single probe. The experiment manipulated the likelihood that a negative probe (not present in current memory set) was part of the memory set of the previous trial. fMRI scans exhibited transient activity in the anterior cingulate, a marker of reactive control, following recent negative trials, but only when these trials were rare. In contrast, when recent negatives were common fMRI scans exhibited sustained pre-frontal activity prior to probe onset, suggesting the use of proactive control (Burgess & Braver, 2010).

There are parallels between control in the Burgess "recent negatives" task and the n-back. Like in Burgess' study, we should expect n-back sequences with frequent lures to bias subjects toward proactive control. On the other hand we should expect sequences that lack frequent lures will provide incentive for a primarily reactive control strategy.

The n-back is a particularly vexing paradigm for the dual mechanisms theory because it involves an unusual interaction of proactive and reactive control. Burgess et al discussed the difficulty in analyzing control strategies, remarking that accurate performance on lure trials requires reactive resolution of conflict, but that this resolution requires a proactively held list of items (Burgess, Gray, et al., 2011).

Althouh the DMC theory specifies two separate neural pathways responsible for proactive and reactive control, respectively, it has not been sufficiently tested in paradigms that may involve both types of control acting simultaneously. The n-back may be such a task, as proactive rehearsals are both the cause of and solution to lure interference. Unlike the AX-CPT, in which proactive and reactive strategies are well separated, the n-back seems to require a combination of proactive and reactive elements.

1.3.5 *n*-back Control Strategies

Success at resolving interference in the n-back depends on the type of control used. Rehearsal of the past n items helps to increase the familiarity of target items, but active maintenance of previous items may increase the interference caused by lure trials (Szmalec et al., 2011; Burgess, Gray, et al., 2011). Effective proactive control, then, would involve maintaining only the nth back item (e.g. "If its an 'A' then press the match button"), relying on episodic memory retrievals to maintain the correct sequence of upcoming target letters on each trial. Evidence from the literature argues against the existence of this method. Performance on n-1 lure trials tends to be worse than that on n+1 trials. In other words, rehearsal of the more recent items (n-1) interferes with performance more so than items no longer being rehearsed (n+1). These results are suggestive of a Baddeley style phonological loop which retrieves and rehearses items in sequence, but is relatively agnostic as to the serial position of any given item. A more effective form of control would be to use a small amount of rehearsal to keep the items in their correct serial order, and add a large amount of activation to the target item. It is possible that this type of control could be learned over time and may explain the substantial training effect that has been found in multiple studies (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi, Buschkuehl, Perrig, & Meier, 2010; Redick et al., 2012). The discovery of the transfer of a general control strategy (rehearsal) to a task-specific strategy (activate only the nth previous item) would go a long way toward explaining training effects in the n-back and other tasks.

There are several possible explanations for how people engage control in the n-back. Juvina and Taatgen modeled sub-vocalization based rehearsal (Juvina & Taatgen, 2007), consistent with Baddeley's original phonological loop. Juvina and Taatgen used the ACT-R architecture which allows only one item to be held in its memory retrieval buffer (localized to the pre-frontal cortex). Thus in order for ACT-R to maintain multiple items it must rehearse by sequentially retrieving each item from declarative memory. For the purpose of this dissertation I will refer to this strategy as "Rolling." The rolling strategy is depicted in the top panel of figure 1.1. Here each new item is compared to the first item in a 3-letter memorized list. On each trial the first item is discarded and the new item is added to the end of the list.

Neurological evidence suggests that proactive maintenance involves a frontal-parietal-straital network which allows for the maintenance of multiple items and their serial order (Owen et al., 2005; Chatham et al., 2011). Chatham's model, based on the connectionist Pre-frontal Basel Ganglia Working Memory architecture (Hazy, Frank, & O'Reilly, 2006; Hazy, Frank, & O'Reilly, 2010), assumes an ability to simultaneously store multiple items in PFC. The model does not actually perform retrievals from an episodic memory store, so it is does not predict that interference from past items would interfere with the current stimulus.

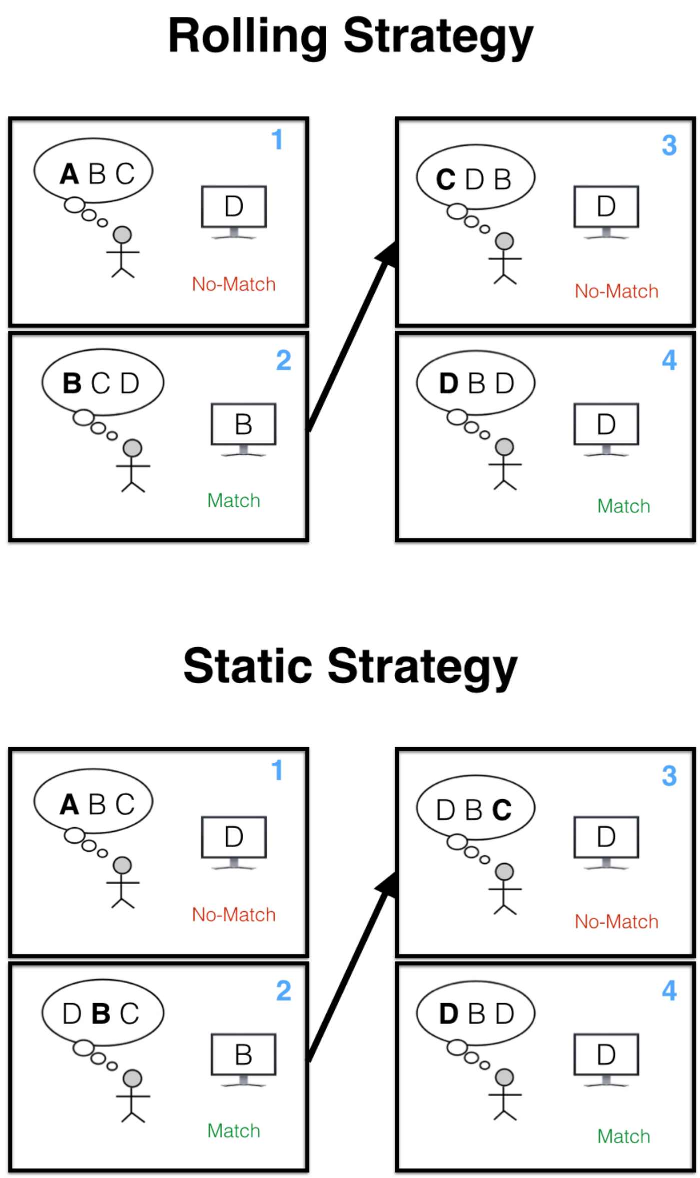


Figure 1.1: Depictions of the rolling and static strategies as applied to an identical sequence of letters (A B C D B D D). Top: The rolling strategy involves comparing new letters to the first item in the list, removing the first item and adding the new item to the end of the list. Bottom: The static strategy involves comparing the new item to the focused item (**Bold**), replacing that item with the new letter and advancing focus to the next item in the list. Of particular interest is the transition between panels 2 and 3, in which the identity of the items remains the same and only the focus changes.

Memory items in Chatham's model are held in pre-frontal cortex while serial position information is updated separately in the parietal cortex. As a result, only the current item needs to be updated on each trial. This strategy, which I will refer to as "Static," is depicted in the bottom panel of figure 1.1. For instance, if the initial three items are (A B C), the serial position marker in parietal cortex would point to the first item (A). If the next item is D, the model would respond "no match" (D != A), and replace the A with the D. The resulting list would be (D B C), and the serial position marker would move to the 2nd position (B). Thus instead of updating the serial position of each item on each trial, the model only updates the "highlighted" item and the current position. As a result, the model dramatically reduces the amount of interference from previous rehearsals.

This static strategy has several advantages over a more conventional rolling rehearsal strategy in which people add new item to the end of the list and advance each item one position. First, the static strategy only replaces one item per trial instead of updating the serial positions of each of the past three items. Second, the static strategy requires no updating on target trials. Finally, because serial position is held separately from the identity of each item, there is less of a chance that an error would occur from retrieving the wrong serial position connected to a given item.

These advantages make the static strategy easier to perform and less likely to lead to predictable errors on lure trials. Because only one item is updated, people using this strategy have more time between trials to rehearse the current items than people using rolling rehearsals. The difference in the updating process is largest after target trials. Consider the sequence (A B C A). After the current letter (A), The memorized list using the rolling strategy changes from ABC to BCA, so in addition to moving each item up the list, the new item added to the end of the list is the same as the item that was just dropped. It is easy to see how this could cause confusion (especially with fatigue). By contrast, the static rehearsal strategy requires no updating. The memory list after the current A is still ABC, and the subject only needs to move his focus of attention to the second letter (B).

In relation to DMC, the static strategy requires proactive control while the rolling strategy does not. In order for the static strategy to function, people must actively maintain the currently focused serial position. The rolling strategy, by contrast, always attempts to match the first item in the list, so subjects can use reactive control, retrieving the last three items in response to each new item.

1.4 Summary

The n-back provides a test case for many theories of working memory and cognitive control. Baddeley's phonological loop explains how the mind can keep multiple items active, but it is not clear how sequential order is maintained. n-back performance arises from the combination of proactive and reactive control processes, making it a natural task to consider from a dual mechanisms perspective. Yet the n-back is a challenge for the DMC theory because unlike many of the tasks commonly used to study DMC, successful performance in the n-back requires simultaneous use of both reactive and proactive control. Because both control types are needed, it is not clear how environmental factors such as the concentration of lure trials and the relative frequency of stimuli will affect control in the n-back. In light of this, it is imperative that we consider the biases caused by the trial sequences we present to our subjects. In the next chapter I will describe the statistical biases found in previous n-back research and discuss how these may affect n-back control strategies.

Chapter 2
n-back: an Expose

2.1 Statistical Bias

Chapter 1 introduced factors that can affect the control strategy people use to perform the n-back. The current chapter details specific examples of trial type distributions (e.g. targets, lures, distractors) and stimulus frequency distributions commonly found in the n-back literature. Sequences common in the literature include many more targets than would be expected by chance and use very few (if any) lures (Chen, Mitra, & Schlaghecken, 2008; Brouwer et al., 2012; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). These sequences also feature strongly skewed stimulus frequencies such that one or two of the possible stimuli are presented at a much higher rate than would be expected by chance. The experiments proposed in this work investigate how these target heavy ratios and skewed stimulus frequency distributions affect the selection of control processes used in the n-back.

Sequences with many more targets than lures increase the value of a reactive activation based strategy. As discussed in chapter 1, interference occurs only on lure trials, when the increased activation caused by a recent item conflicts with a controlled recollection of the third previous item. When targets vastly outnumber lures, recently seen letters are highly likely to be targets, so a strategy in which responses are determined solely by the activation of the current letter in memory will be very successful. Sequences in which lures outnumber targets, by contrast, diminish the value of memory activation. In these sequences recently seen items are more likely to be lures than targets, so responding only to memory activation without proactive maintenance of the target item will lead to numerous errors.

Skewed stimulus frequency distributions also increase the value of a reactive strategy by increasing the activation of items most likely to be targets. As I will discuss in detail in this chapter, letters presented most frequently have a higher conditional probability of being a match than those presented less frequently. A strategy which emphasizes memory activation will naturally lead to responses of “match” for high frequency letters and "no-match" for low frequency letters. Because high frequency letters are conditionally more likely to be matches, this strategy has a high rate of success.

Variations of target to lure ratios and stimulus frequency distributions in the design of n-back studies make it problematic for researchers to claim the n-back as a pure measure of working memory. Without understanding the n-back's strategic puzzle, it is impossible to make accurate conclusions regarding specific cognitive functions that make up those strategies.

2.1.1 Definitions

This section will introduce and define the terminology used throughout the remainder of the work. The hypotheses of the work relate to the ratio of targets to lures and the stimulus frequency distribution of the experimental sequence. Terms I will use to denote differences in these statistical properties of sequences are as follows:

1. *T:L Ratio* - The ratio of targets to lures
   (a) *Low T:L ratio* - Target to lure ratios of less than 1:1 (e.g. 1:2 or 2:3), such that lures outnumber targets. These ratios will also be referred to as *Lure Heavy Ratios*.
   (b) *High T:L ratio* - Target to lure ratios of greater than 1:1 (e.g 2:1 or 3:2), such that targets outnumber lures. These ratios may also be referred to as *Target Heavy Ratios*.
2. *Stimulus Distribution* - The distribution of the number of occurrences for each possible stimulus item in a given sequence.
   (a) *Skewed Sequences* - Sequences in which the some stimuli are presented much more often than others. For the purpose of the proposed experiments using 8 possible items, skewed sequences are those in which the four most commonly presented items are used for at least 2/3 of the trials in a sequence. A skewed sequence of 64 trials selected from the first 8

letters of the alphabet might contain 18 A's, 13 B's, 8 C's, 8 D's, 5 E's, 5 F's, 4 G's and 3 H's. Sample skewed sequences can be found in appendix A.

(b) *Even Sequences* - Sequences in which each stimuli is presented the same number of times in a sequence. Thus for a 64 trial sequence, each of 8 stimulus items would be presented exactly 8 times. Sample even sequences can be found in appendix A.

(c) *Lumpiness* - The amount of repetitions over a short span of trials. Lumpiness varies within a sequence, regardless of the overall distribution of stimulus frequencies throughout the sequence. For instance, the subsequence (A, B, C, C, B, C ,B) is more lumpy than the sequence (A, B, A, D, F, B, C).

3. *n-back specific terminology*

   (a) *3-back/nth back item* - The third previous item presented. For the purpose of this study, the nth back item is the third previous item. These trials will be referred to as targets or matches throughout this work.

   (b) *Match/Target* - The terms match and target are used interchangeably to refer a stimulus that matches the 3-back item.

   (c) *n-1 item* - A lure trial matching the stimulus presented n-1 trials in the past. For the purpose of this study, this is the 2nd previous item.

   (d) *n+1 item* - A lure trial matching the stimulus presented n+1 trials in the past. For the purpose of this study, this is the 4th previous item.

4. *Other Terminology*

   (a) *Random Sampling* - The practice of choosing an item from a pool of available items via a pseudo-random computer algorithm.

   (b) *Random Sampling With Constraints* - Skewed sequences are constructed by randomly sampling each item from a pool of letters that satisfy a pre-determined trial type. For instance to select a non-lure distractor following the sequence A, H, B, D, we would sample from the subset of letters (C, D, E, F, G) that would not create a target or lure trial. Thus these sequences are constructed via random sampling with constraints. Even sequences also use random sampling with constraints, but the constraints are much more detailed. A full explanation will be provided later.

(c) *Expected Distribution* - The distribution created from the average of multiple iterations of random selection. For instance, over thousands of trials we expect 50% of coin flips to land heads. Likewise, the expected distribution of non-constrained n-back sequences should be even (e.g. 8 of each of 8 items in a 64 trial sequence) because that is what the average distribution should be given a large number of iterations.

It is important to emphasize the difference between expected values and random sampling. As Abelson remarked in his 2012 book, "Chance is lumpy" (Abelson, 2012). Random sequences are skewed, but the average of multiple iterations of random sequences converges to an even distribution. In this chapter I will demonstrate that while researchers may believe they are using expected distributions (even) because they are sampling randomly, they are actually presenting their subjects with heavily skewed sequences.

2.1.2 Target:Lure Ratio

The relative proportions of targets, lures, and distractors is critical in determining optimal control strategies. With random selection, expected distribution of stimuli yields a rate of target presentation that is inversely proportional to the size of the stimulus set. For instance if chosen from a set of 8 possible stimuli, we would expect approximately 1/8 or 12.5% of trials to be targets. With 20 possible stimuli, we would expect 1/20 or 5% matches. Lures should be more common than targets, as the expected distribution would yield slightly less than twice as many lures as targets. This creates an experimental design problem, as the most interesting trials (targets and lures) become exceedingly rare. Consider an extreme example. If we randomly select a 64 item sequence from a pool of 20 available stimuli, we expect to choose a mean of 3.2 matches and 6.4 lures (assuming none of the lures and matches co-occur). Thus 9.6 out of 64 (15%) trials would be targets or lures. Conversely, subjects could correctly reject 85% of trials simply by recognizing that they have not been presented for a substantial amount of time.

In order to study the ability to maintain items in memory and thus recognize matches, researchers tend to increase the rate of target presentation. Most n-back studies use the methodology of Braver, Cohen, et al., 1997 and Awh et al., 1996 and present targets on approximately 33% of trials. Depending on the size of the stimulus set, this represents a significant deviation from the expected distribution. In the paragraph above I presented an example of a sequence created by random selection. But if we constrain the selection process such that 33% of trials are matches, we would instead present 21 targets, 4 lures, and 39 distractors. We have selected 700% more targets, and actually decreased the prevalence of lures!

In addition to increasing the number of targets presented, researchers have unintentionally made those targets easier to detect. Taken together, targets and lures represent the trials most likely to be matches. For the purpose of this thought experiment I will refer to these as "recently presented items." In constrained sequence described above, 21 of these 25 recently presented items are targets. Thus matches occur on 84% of recently seen trials. By contrast, the non manipulated version (3 targets, 6 lures, 55 distractors) presented targets on only 3 out of 8 recently presented trials, making matches harder to differentiate from lures.

Table 2.1 provides an analysis of the target and lure proportions in a selection of previous n-back studies. While this is not an exhaustive list, it represents the approximate range of target:lure (T:L) ratios found in the literature. It is not common for lures to outnumber targets. Indeed, most researchers use T:L ratios such that there are many times more targets than lures. Despite dramatic deviance from the frequencies of these trial types we expect from non manipulated sequences, virtually no discussion is found in the literature as to how these methodological issues might affect performance.

Manipulation of trial type frequencies can affect the type of control used to

Table 2.1: Statistical Properties of n-back studies

Study	Stimulus Pool	Targets	Lures	Distractors	T:L Ratio
Chen 2008	20	20	4	40	5:1
Braver 1997	20	8	?	?	
Nystrom 2000	18	6	2	11	3:1
Brouwer 2012	21	16	3	29	5:1
Dieber 2007	20	7	2	21	7:2
Harbison 2011	?	5	5	15	1:1
Jaeggi, 2008	8	6	3	11	2:1
Juvina 2007	Varied	10	5	Varied	2:1
Juvina 2007 exp 2	Varied	10	10	Varied	1:1
Kane 2007	8	8	8	32	1:1
Schmiedek 2009	16	12	12	12	1:1
Szmalec 2011 exp 1	20	15	6	24	5:2
Szmalec 2011 exp 4	20	12.5	9.5	23	1.32/1
Tsuchida 2009	26	20	0	100	20:0

perform the n-back. The DMC literature suggests that a high T:L ratio should bias people toward reactive control. A high T:L ratio means that most recently presented trials will be matches, increasing the value of memory activation (T. Braver, personal communication). A low T:L ratio, on the other hand, should bias people toward a proactive strategy, as most recently presented trials will be lures, not targets. Thus proactive maintenance of the 3-back item is required to disambiguate target and lure trials .

Although the effect of T:L ratio has not been intentionally manipulated in any study of which I am aware, Juvina and Taatgen noted the phenomena in their 2007 study (Juvina & Taatgen, 2007). Juvina presented 10 targets and 5 lures per block and found an unexpected negative correlation between performance on targets and lures, respectively.

> Unexpectedly, the correlation between correctness on targets and correctness on foils [lures] was negative (r^2 = -0.53, p = 0.0004). Participants tended to score either high on targets and low on foils [lures] or vice-versa. This is an indication that some of the participants mani-

> fested what we called a "react-to-repetition" effect: they were tempted to react to a repeated item regardless whether they knew or not that it was a target or a foil. Since the number of targets was higher than the number of foils such a strategy would pay off overall. Other participants, who scored low on targets, scored high on foils because non-reaction to foils counted as correct answer. In both cases the correctness score was artificially increased. (Juvina & Taatgen, 2007, p.74)

These results are consistent with a dual mechanisms account which would expect people to use reactive strategies in response to a high (2:1) T:L ratio. Reactive control would lead to a high number of match responses to recently presented items, resulting in a dissociation between accuracy on target and lure trials. Such a response bias should lead to high accuracy on target trials and low accuracy on lure trials, and this is exactly what happened.

As a result of these insights, Juvina and Taatgen conducted a second experiment which featured equal frequencies of target and lure trials. In this case the correlation between accuracy on targets and lures was positive. Subjects who were accurate in recognizing targets were also accurate in rejecting lures. Likewise, subjects who struggled to detect targets also performed poorly on lure trials. Such an outcome is consistent with an expectation that the equal likelihood of targets and lures lead subjects to adopt proactive control strategies.

It may be confusing that some subjects in this study performed poorly on both targets and lures. The authors do not go into detail, but it is an important point. The inability to either accurately identify matches or reject lures suggests the use of a particularly ineffective rehearsal strategy. Such a strategy might use rehearsal to raise the activation of the previous 3 items but do a poor job of maintaining position information. Thus lures would be difficult to reject because they would be highly active, but targets would be equally difficult to identify due to a lack of correct position information. Still, this would be classified as a proactive strategy. A reactive strategy would have resulted in results similar to those found in Juvina and Taatgen's first experiment, with accuracy rates dissociated by trial type. These poorly performing subjects highlight an interesting point sometimes missed when discussing DMC. Proactive strategies are not always superior to reactive strategies. For these subjects the choice of control strategy was determined by the T:L ratio, not by the success of their chosen strategy.

It is likely that subjects in other studies have employed reactive strategies in response to high T:L ratios. However, most n-back researchers do not distinguish n-1 and n+1 trials as lures and sum across all trials, reporting only total accuracy. Thus it is impossible to determine if this effect has been replicated in the literature. An exception is Harbison et al., 2011, which like Juvina and Taatgen's experiment 1 showed a dissociation between accuracy on target and lure trials. Although Harbison did not report correlation statistics, he did report that subjects performed much better on lures (∼80%) than targets (∼60%). Interestingly, Harbison used a 1:1 target:lure ratio and still found a dissociation of target and lure performance that suggested a bias toward "no-match" responses. One possible explanation is that Harbison's study used an adaptive system much like that used by Jaeggi (2008) which emphasized overall accuracy (including distractors), so subjects may have had an incentive to respond "no-match" to the vast majority of trials (distractors and lures) that did not match.

Kane, Conway and Miura used a 1:1 target:lure ratio and like Juvina and Taatgen found no response bias in a 3-back task (Kane et al., 2007). Kane's study carefully counterbalanced all stimuli such that each stimulus appeared an equal number of times. As will be discussed in the next section, this may have further biased subjects toward proactive control. Given the results of the studies by Juvina and Taatgen, Kane et al, and Harbison, it seems that the target to lure ratio is one of many task factors that affect control strategy selection. Juvina and Taatgen demonstrated that all else being equal, a manipulation of target to lure ratio can change the control bias. But Harbison and Kane demonstrated that other task factors, including stimulus frequency distributions and task instructions, may also affect control. In the next section I will discuss how stimulus frequency distributions may affect control in the n-back.

2.1.3 Selection Frequencies

Manipulations of trial type frequencies also create unintended biases in the choice of stimuli. Given x possible stimulus items and m trials per block, each stimuli should be presented approximately m/x times per block. For instance, a 64 trial sequence with 8 possible stimuli should include 8 occurrences of each item. However, manipulations of trial type frequencies common in the literature lead to strongly skewed stimulus frequencies. To demonstrate, I simulated likely trial sequences given constraints common in the literature.

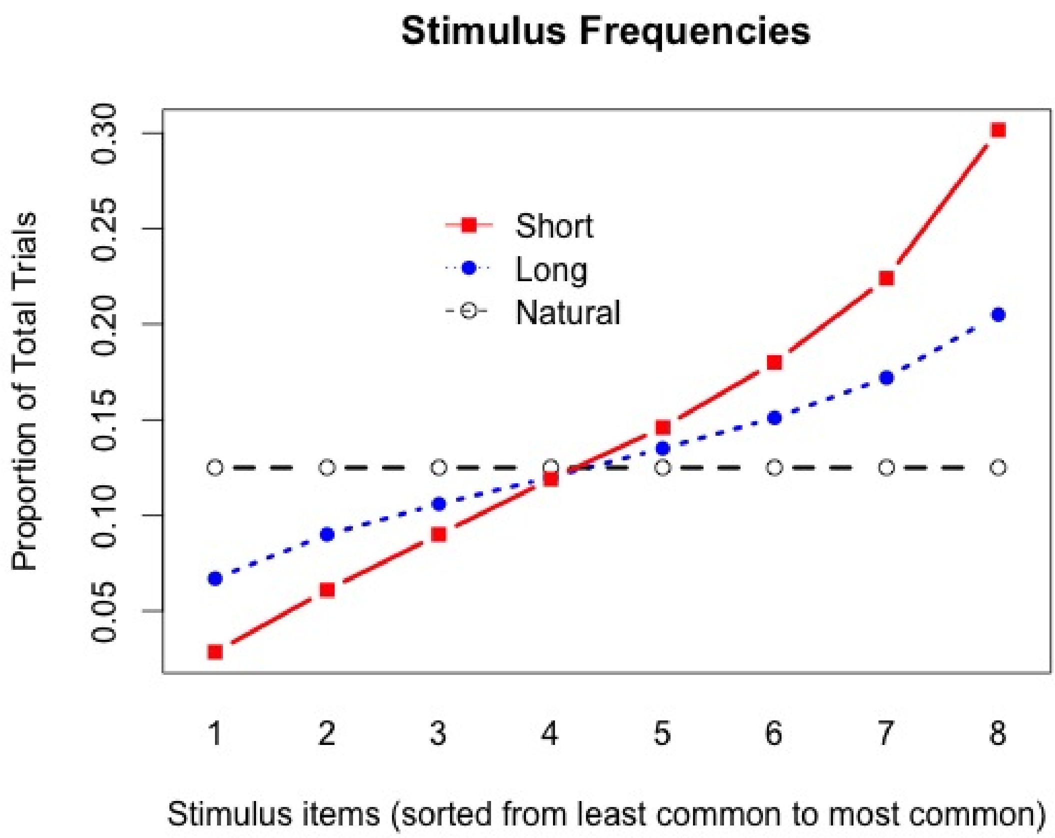


Figure 2.1: Simulated stimulus frequencies based on the methodology of previous studies. Each of eight stimuli were sorted from least common to most common. The natural line represents the default expectation that with random sampling each item in the stimulus set will be used approximately equally often. Four our 8 item list that means on 12.5% of the trials. The "Short" line represents mean values for a 24 trial block. The Long line represents values for a 67 trial block.

Figure 2.1 presents mean stimulus frequencies calculated from 1000 simulations for two example 3-back sequences. The first example (short) is constrained to include 5 target trials and 4 lures out of 24 trials. The second example (long) includes 16 targets and 8 lures out of 67 trials. The horizontal line represents what would be expected if each of the n stimuli had an equal probability of being sampled at any point in the sequence. In each case, design manipulations strongly affect the distribution of stimuli.

In general, random selection of n-back sequences follows Abelson's first law, "Chance is Lumpy" (Abelson, 2012). Manipulation of trial type frequencies in a randomly selected sequence results in even more lumpiness. Specifically, large differences between the frequencies of commonly selected and rarely chosen items are common. This effect is amplified for short sequences. Note that the despite having a relatively even 5:4 target to lure ratio, the 24 trial sequence selects the 2 most common items over 50% of the time.

In many cases the 2 to 3 most frequent stimuli make up the vast majority of trials, far more than would be presented in an even sequence. In these sequences the vast majority of the target and lure trials are concentrated at the top of the stimulus distribution. For instance, the long sequence in figure 2.1 reflects a trial sequence where 30 of 64 trials (46%) would feature one of the two most common items. These two stimuli would have been presented on an even higher proportion of target and lure trials, a mean of 13 of 16 target trials (81%) and 5 of 8 lure trials (62.5%). Since these familiar items tend to be targets at a higher rate than other items, a reactive, familiarity based strategy can be quite successful.

2.2 Summary

The n-back task provides fertile ground for research into how statistical manipulation affects strategic control. This chapter detailed the unintentional manipulation of two important properties of stimulus sequences, namely the T:L ratio and stimulus frequency distribution. The experiments detailed in this work manipulate these factors in an attempt to determine how the mind adjusts control processes to match the environment.

In the next chapter I will introduce a novel procedure used to create unbiased n-back sequences, discuss possible behavioral markers of control processes, and propose hypotheses predicting how n-back performance will vary as a function of the T:L ratio and stimulus distribution of stimulus sequences.

Chapter 3
Measures and Hypotheses

3.1 Sequence Generation

The current work aims to determine the control dynamics of the n-back by investigating two features of n-back sequences, the T:L ratio and stimulus frequency distributions, respectively. The statistical properties of the n-back make these sequence factors difficult to dissociate. As discussed in the previous chapter, manipulation of T:L ratios tend to cause unintended manipulation of stimulus frequencies. In order to dissociate the effects of T:L ratio and stimulus frequency distributions, we must be able to vary the relative frequency of target and lure trials without disrupting global stimulus expectancy.

A discrete optimization algorithm (Barry & Ralph, 2014 in preparation) allows for the creation of even sequences which manipulate T:L ratios without altering the stimulus frequency distribution. The algorithm manipulates strings of trial type dependencies to arrange items in a way that preserves an even distribution of stimulus occurrence while satisfying pre-determined T:L ratios. Examples of these even sequences are included in the appendix. Each sequence includes 8 observations each of 8 possible stimuli. Thus on any given trial the likelihood of each possible item is equivalent. Every item is also approximately equally likely to be a target, lure, or distractor. Thus the stimulus sequence provides no statistical help in predicting whether a given item is likely to be used as a match trial.

The experiments included in this dissertation use 4 types of sequences, manipulated between subjects. The 4 sequence types are skewed-24, skewed-8, even-24, and even-8, named by the stimulus distribution and the number of targets included. Each block of trials is 64 trials long, not including 3 initial no-response trials. 24 target sequences included 24 targets, 8 lures, and 32 non-lure distractor trials. 8 target sequences included 8 targets, 24 lures, and 32 non-lure distractors. The even sequences were compiled with the discrete optimization algorithm, and the skewed sequences were compiled using a more conventional approach in which trial types (targets, lures, distractors) were randomly arranged and stimuli that satisfied the current trial type were randomly selected. For instance, if the previous 4 items were A B C D, and the next trial-type chosen was a distractor, then a letter from the subset [D, E, F, G, H] would be randomly selected. A, B, and C would be removed from the possible list because A and C would create a lure trial and B would create a match. The python code used to construct these lists can be found in the appendix.

In order to remove some of the variability of stimulus distributions between skewed sequences, I chose a subset of compiled skewed sequences which included similar distributions. Namely, I selected sequences in which the 3 most common stimuli made up at least 40 of the 64 trials. This selection is justified because the even sequences have no variability (they all have 8 of each item), so it made sense for a comparison group of sequences to also have as little variability as possible.

3.2 Local Lumpiness

It is important to note that although all letters are presented the same number of times, even sequences are not uniform in regards to local stimulus repetitions. The term uniform would suggest that you could divide a sequence into several subsequences in which letters are evenly represented. For instance, the sequence (A B C D B D C A D A B C C D B A) is uniform in that you could break it into four 4-letter subsequences in which each letter is presented once. The even sequences used in the current study do not follow this pattern.

Even sequences preserve some of the local lumpiness of the skewed n-back sequences used in the literature, but differ because the same few letters are not repeated over and over throughout a sequence. In order to quantify local lumpiness, I used the number of repetitions which occurred in the previous 8 stimuli. For instance, if the previous 8 items were A B C B C A D H, the current lumpiness of the sequence would be 3 because 3 of the eight stimuli (A, B, and C) were each repeated once among the 8 trials. A perfectly uniform (non-lumpy) subsequence (e.g. A B C D E F G H) would yield a lumpiness score of 0 (no repetitions).

An analysis of the 1724 sequences used in the third experiment found that the skewed sequences (mean lumpiness = 3.3, SD = .24) were significantly more lumpy than even sequences (mean lumpiness = 2.78, SD = .19), $t(1640) = 49.46$, p <.0001, d = 2.41. Although even sequences are less lumpy than skewed sequences, their lumpiness does vary considerably within sequences. Figure 3.1 depicts histograms of within-sequence lumpiness for skewed and even sequences, respectively. Although approximately 40% of trials of both sequence types follow 8 item subsequences with 5 unique letters (3 repetitions), the shape of the distributions differ. Nearly 40% of trials in skewed sequences follow subsequences with 4 or fewer unique letters, and 40% of trials in even sequences follow subsequences with 6 or more unique letters.

Despite these differences, lumpiness variation within sequences makes it possible to analyze how people adapt to trial-by-trial changes in the task environment. Specifically, when using reactive control people will be sensitive to changes in lumpiness, but dedicated use of proactive control should lead to static performance regardless of the current state of the sequence.

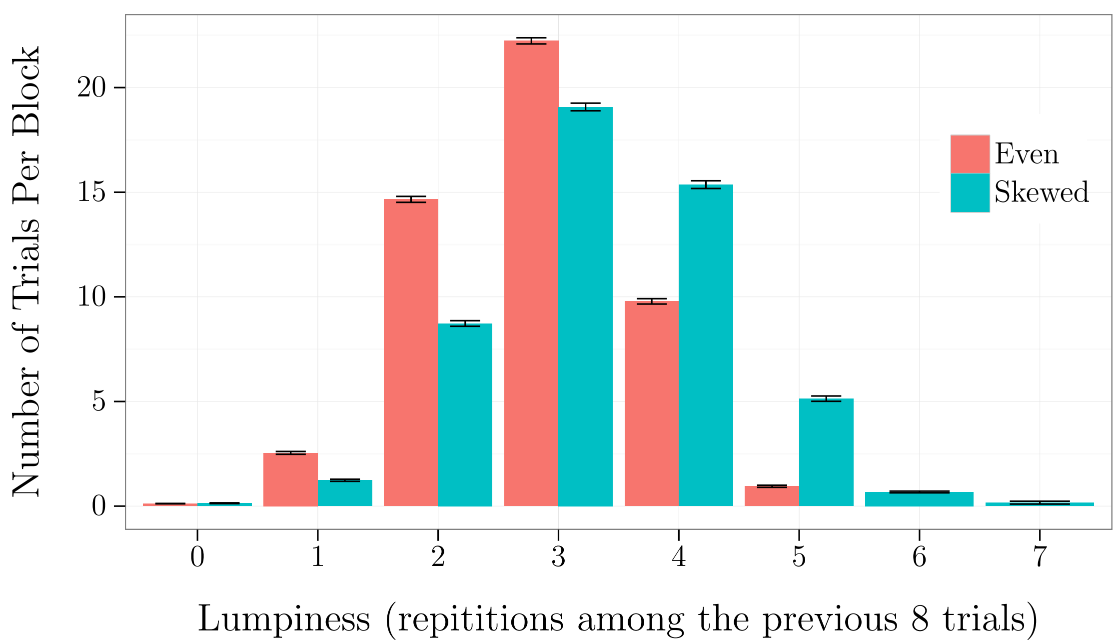


Figure 3.1: Histogram representing the number of trials per block following various configurations of past letters. The x-axis is the number of stimulus repetitions in the previous 8 trials.

3.3 Measures

In this section I will discuss possible methods for differentiating between control modes using only behavioral (e.g. response time and accuracy) measures. Researchers studying DMC generally rely on fMRI scans to differentiate between proactive and reactive control. Proactive control features sustained activity in DLPFC while reactive control leads to transient activity in other areas of pre-frontal cortex and medial frontal cortex immediately following the presentation of lure trials (Gray, Chabris, & Braver, 2003; Braver, Gray, & Burgess, 2007; Burgess, Gray, et al., 2011). As the current study does not utilize fMRI technology, I will instead use only behavioral analyses to differentiate between control states.

3.3.1 Behavioral Measures

Most n-back studies in the literature employ only the most basic of measures, limited to mean response time and accuracy. Some (Szmalec et al., 2011; Juvina & Taatgen, 2007; Harbison et al., 2011) separate results by trial type (e.g. targets, lures, and distractors), but even these are insufficient to determine how strategic variation affects performance. A more detailed analysis comparing performance on different trial types and trial by trial changes in performance will provide information relevant to the hypotheses. The nature of proactive and reactive control, respectively, will lead to different predictions of these more detailed analyses.

Researchers studying memory recognition commonly use Signal Detection Theory (SDT), a method for determining response bias and comparing performance to chance. SDT compares rates of correct hits (responding "yes" on a matching trial) to false alarm errors (responding "yes" on a non-matching trial). Two SDT measures were used to analyze data in the current work, Sensitivity (d') measures the relative ability to correctly identify targets and avoid false alarms on no-match trials. Criterion (c) relates to a bias to respond yes or no.

SDT provides an explanation for Juvina's discovery that performance on targets and lures diverged in his experiment 1 (Juvina & Taatgen, 2007). In the extreme case, if a subject answered yes to all familiar items, she would be correct on nearly all targets and incorrect on all lures. A SDT analysis of this data would yield a d' of 0 (no difference from chance). The same d'=0 would be recorded if a subjects answered no to all trials, reflecting perfect performance on lure trials but no targets correctly identified.

Although the two extreme examples discussed above would yield identical d' values they clearly represent different strategies. A second SDT measure, criterion (c) helps to capture this difference. The Criterion value is the level of certainty a subject must have before he will respond "yes". If a subjects is biased to answer yes to all familiar trials, then his criterion value will be very low. If he is biased to answer no, then he will have a very high criterion.

Although SDT has been used previously to measure n-back performance (Kane et al., 2007), the statistical variation of n-back sequences makes SDT measurements problematic. SDT determines how likely it is that the subject is able to recognize a signal in a noisy environment. In the n-back, the activation level of the current letter in memory likely causes variation in the noise distribution. False alarms are more likely to occur on lure trials than non-lure distractors, and the chance of an error decreases as the time between identical presentations increases.

A potential solution is to focus only on match and lure trials, eliminating distractors from the analysis. Lure and match trials have similar lag, each being presented between two and four trials in the past. Also, lure trials are qualitatively different from other distractors because of the rehearsal processes thought to add activation (and resulting noise) above what would occur without intervention. Unfortunately, this method eliminates the ability to analyze the effect of lag on accuracy, which might help differentiate between control modes. Reactive strategies lead to a decrease in error rates with increasing lag, so a relevant measure of sensitivity should include all trials. For the current work I used all trials to calculate measures of sensitivity and bias with the caveat that d' values will be generally elevated due to the inclusion of distractors (which have very low false alarm rates).

Although researchers interested in memory tend to focus on accuracy rates, response times are more informative when attempting to isolate control mechanisms and strategies. Studies of cognitive control often focus primarily on response time as the behavioral measure of choice (Yeung, Bogacz, Holroyd, & Cohen, 2004; Yeung, Ralph, & Nieuwenhuis, 2007). In fact researchers in these studies specifically instruct subjects to attempt to achieve a constant error rate in order to remove response time variability caused by the speed accuracy trade-off.

I analyzed response times in several ways. First, I analyzed response time as a function of the lag since the last identical presentation of the current stimulus. Proactive control results in the preparation for the target letter, so response times on match trials should be significantly faster than those on lure and distractor trials. Response times for lure trials should be slower than other distractors because rehearsal leads to increase in activation of lure letters (Szmalec et al., 2011). Reactive control should result in an inverse linear relationship between response time and lag such that response times decrease as lag increases. Because control in the n-back is a complex combination of reactive and proactive control, I expected to see a mixture of these control profiles. Nonetheless, comparisons of the difference of response time between trial types, and the linear relationship between response time and lag will allow me to make conclusions regarding control mode selection.

I will also attempt an analog of the difference based analysis common in the cognitive control literature. Task switching studies, for instance, often measure the switch cost, the difference in response times between switch trials and repeat trials. Likewise, flanker and Stroop researchers often measure the congruency effect, the difference between response times on incongruent and congruent trials.

In the current work I measured the difference between response times on lures and targets (separately for n-1 and n+1 lures). The rationale is that target trials represent congruency in the n-back (in which a familiar item is also the 3rd previous letter) while lure trials represent the most severe form of incongruency (highly familiar items that are not the 3rd previous item). Data will be separated by lure type because n-1 and n+1 are qualitatively different trials. n-1 trials are in the current rehearsal window, so excessive rehearsal should lead to more conflict between activation and recollection and result in larger differences between match and lure trials. An n+1 lure, by contrast, is not in the current rehearsal window, but it was the target item on the previous trial. Proactive strategies such as the static strategy which provide extra activation to the target item will lead to higher activation of

n+1 lures and result in slower response times.

A final analysis focuses on the relationship between response time and local lumpiness. Lumpiness should affect response times more strongly for subjects who adjust control strategies to the environment. Subjects open to reactive influences will likely adjust the amount of rehearsal they engage in to changes in the lumpiness of the stimulus sequence. On the other hand, subjects focused on maintaining and remembering the last three items will show no effect of lumpiness. In theory, the rehearsals used to maintain the previous three items should raise the activation levels of these letters well beyond the background noise which will fluctuate with varying degrees of sequence lumpiness. By contrast, reactive strategies will take advantage of changes in lumpiness to adjust strategies within sequences. Response times for these subjects will show a linear relationship between response time and lumpiness. This effect should be most pronounced on n-1 lure trials, which are most sensitive to changes in rehearsal rate.

3.4 Statistics

Statistical analysis for all experiments was performed with the R statistical programming language (R Core Team, 2013). Although each study was designed with an equivalent number of subjects per condition, some analyses focused on self-reported strategy choices which were unbalanced. For consistency, analysis of variance were performed using the ezANOVA (Lawrence, 2011) function using type III sum of squares. For balanced designs, Type III sum of squares are identical to the type I sum of squares used by the base aov function in R but type I sum of squares are not suggested for unbalanced designs. Please refer to the appendix as indicated in the results sections for full ANOVA tables for each analysis detailed.

3.4.1 Effect Sizes

Effect sizes are reported in the form of partial eta-squared (η_p^2) for each significant and marginal effect. Where not noted, per Cohen's original criteria, η_p^2 values between .01 and .05 are considered to be small effects, between .06 and .14 are considered to be medium sized, and larger than .14 are considered to be large effect sizes.

3.5 Hypotheses

The general hypothesis of this work is that people develop strategies which employ a combination of proactive and reactive control mechanisms, and the statistical properties of the stimulus sequence bias processing toward one or the other. It is important to note, as discussed in chapter 1, the use of proactive and reactive control may be seen as a separate issue from the choice of a specific memory strategy. That is, people may use a rolling memory strategy, but differ in the number of rehearsal iterations they perform between stimuli or the amount of specific rehearsal of the target item. Likewise, the use of the static strategy may vary in the amount of rehearsal and emphasis provided by the pointer. Here I will delineate theoretical hypotheses relating to the statistical properties of n-back sequences. Table 3.1 provides a reference for the specific hypotheses, which are described in detail below.

H1: Effect of Target Ratios. T:L ratios create a strategic bias by affecting the likelihood of interference on recently presented stimuli and thus the usefulness of a reactive strategy. Target heavy ratios reduce the likelihood of interference (recently presented stimuli are more likely to be targets) and increase the usefulness of a reactive activation-based strategy. Thus people will be less likely to use proactive rehearsals when encountering target heavy T:L ratios. In contrast, lure heavy T:L ratios will increase the likelihood of interference when encountering a recently seen item and thus increase the need to hold the nth previous item active. Thus subjects presented with low T:L ratios will tend to perform more proactive rehearsals in order to keep the target item highly active.

H2. Effect of Stimulus Distribution. Skewed stimulus sequences increase the activation of a small subset of items which are commonly presented as targets. As such, these sequences will require less rehearsals to accurately recognize matches. I expect that skewed sequences will lead to reactive strategies which take advantage of changes in sequence lumpiness to decrease the frequency of rehearsal. This will be evidenced by a linear relationship between response time and lumpiness. On the other hand, even sequences such as those used in Kane et al., 2007 will require more proactive control to successfully identify matches. Each letter in an even sequence is presented with equal frequency, so without proactive control all stimuli will have equal activation. It unlikely that subjects will be able to distinguish between familiarity caused by frequency and recency. As a result, people presented with even sequences will show no effect of lumpiness. Also, because of differences in control strategies I expect the RT difference between target and lure trials to be greater for even sequences than for skewed sequences.

An alternative hypothesis, articulated by Todd Braver in a personal communication is that skewed sequences will bias people more strongly toward the mode of control determined by other task factors. For instance when lures are more common than targets, skewed sequences would make proactive control even more valuable. But in a reactive-friendly sequence, when targets are more common, skewed sequences would make reactive control even more valuable.

H3: Static vs Rolling Rehearsal: Differences in performance will emerge between subjects using static and rolling strategies. The static strategy provides specialized activation to the target letter, and as a result will outperform the rolling strategy. The rolling strategy will also show more evidence of reactive influences (e.g. susceptibility to n-1 lure interference).

Table 3.1: Hypotheses

T:L Ratio	Distribution	Control Mode Prediction
3:1	Skewed	Reactive
3:1	Even	Mixed
1:3	Skewed	Mixed
1:3	Even	Proactive

Chapter 4
Experiment 1

Experiment 1 investigated how statistical manipulations common in the literature affect control strategies and performance in the n-back task.

4.1 Methods

4.1.1 Participants

80 subjects (43 female) were recruited from a pool of psychology students at Rensselaer Polytechnic Institute and awarded research credits for their participation [1]. Subjects completed 13 64-trial sequences over the course of approximately 90 minutes.

4.1.2 Materials

Experiments were performed on a Apple Mac Mini running the OSX operating system. Experimental software was custom coded in the LISP programming language. Experimentation occurred inside a sound resistant experimental booth. Subjects responded by pressing keys on a standard keyboard. Stimuli (letters) appeared every 3000 ms and remained on-screen for 1000 ms. Figure 4.1 is a schematic of the trial procedure. A small black box appeared in the center of the screen if no response was made within 1500 ms of stimulus onset, signaling to the subject the need to respond more quickly. This box remained on-screen for 1000ms. Stimuli appeared with a 3000 ms SOA.

[1] Six subjects were removed from the analysis due to an error in administering the experiment. These subjects were accidentally presented with a set of sequences which did not conform to the guidelines of the experiment. Removal of these subjects resulted in a slightly unbalanced distribution of subjects to conditions as can be seen in table 4.1.

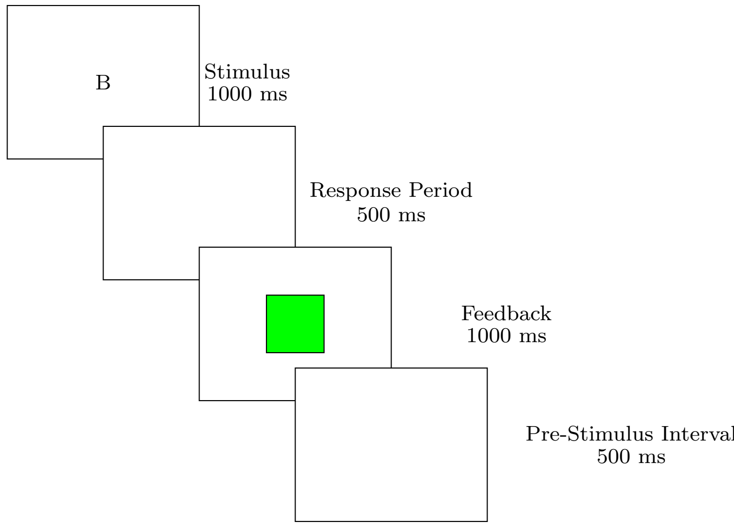


Figure 4.1: Trial process for experiment 1. Stimuli are displayed with a 3000 ms SOA. Stimuli remain on screen for 1000 ms. A colored feedback box (green or red) appears 1500 ms after stimulus onset and remains on-screen for 1000 ms. If no response is recorded within 1500 ms of stimulus onset, a black box appears for 1000 ms, signaling the need to respond more quickly in the future). Experiments 2 and three omitted the feedback boxes. Experiment 2 also omitted the black no-response box.

4.1.3 Design

The design of this experiment was 2 (T:L ratio) x 2 (stimulus distribution) x 12 (block) with T:L ratio and stimulus distribution manipulated between subjects. Following a 64 trial practice block completed with the researcher in the experimental room, each subject completed 12 experimental blocks. Subjects were randomly assigned to one of four between-subjects conditions. The conditions spanned the two T:L ratios (24 matches/8 lures or 8 matches/24 lures) and two stimulus distributions (even or skewed) . Refer to table 4.1 for details regarding these sequence types. 4 sets of 13 64-trial (+ 3 initial letters) sequences were generated ahead of time and each subject in a given condition was presented with the same 13 sequences (one sequence per block and one practice block). Sequences were generated either by discrete optimization (even sequences) or with the python code in the appendix (skewed sequences). These sequences were described in detail in chapter 3.

All sequences sampled from the following 8 phonologically distinct letters (B, F, H, J, M, R, X, Z). These letters were modified from the 8 letters used by Kane et al., 2007 (B, F, H, K, M, R, Q, X). I removed K and Q because they begin with a similar sound and replaced them with J and Z. The resulting list includes 8 letters which begin with different sounds and sample equally from the beginning, middle, and end of the alphabet.

Table 4.1: Sequence Conditions for Experiment 1

Condition	Targets	Lures	Distribution	T:L Ratio	Subjects
Skewed-8	8	24	Skewed	1:3	20
Skewed-24	24	8	Skewed	3:1	19
Even-8	8	24	Even	1:3	17
Even-24	24	8	Even	3:1	18

4.1.4 Procedure

In order to introduce the task, the researcher narrated a short slide-show explaining only that the goal of the task was to determine if the current letter matched the third previous letter. The slide-show contained a short sequence, and the researcher helped the subject to understand the basic rules of the task. Researchers did not discuss strategies (such as rehearsal), so as not to prime subjects to use a proactive strategy. Researchers instructed subjects to respond by pressing a key within 1500 ms of stimulus onset. Subjects pressed the “a” key on a standard keyboard to indicate a match and the “d” key for a non-match. The computer presented each letter in black font color at the center of a white screen. Stimuli subtended 1 degree of visual angle and remained on the screen for 1000 ms, followed by a 500 ms extended response period. Following the response period, a colored feedback box appeared for 1000 ms. Finally, the feedback disappeared and the next stimulus appeared 500 ms later. Thus the total SOA was 3000 ms. See figure 4.1 for the details of trial timing.

After each response the computer provided feedback in the form of a colored box in the center of the screen. A green box denoted a correct response and a red box denoted an error. A black box appeared if no response was made within 1500 ms of stimulus onset. After each block of trials the computer displayed feedback for the previous block of trials (% correct for all trials, % correct for match trials and average response time).

4.2 Results

Statistical analysis for all experiments was performed with the R statistical programming language (R Core Team, 2013). Except where noted analysis of variance was performed using the ezANOVA (Lawrence, 2011) function. All analyses were performed using type III sum of squares. Please refer to the appendix for full ANOVA tables for each analysis detailed.

4.2.1 Response Accuracy

Response accuracy was analyzed using SDT parameters (d' and c), calculated for all trials by block. An analysis of variance was performed for the 2 T:L ratio (Target Heavy / Lure Heavy) x 2 Stimulus distribution (Even / Skewed) x 12 (Block) design. For each analysis I performed an ANOVA and linear trend analysis on the full data set (blocks 1-12). If the analysis showed a significant linear or quadratic trend of block I subsequently eliminated blocks 1-3 and performed an ANOVA on the truncated dataset. In most cases this process removed the trend, so I considered the truncated dataset to reflect "steady state" performance. I calculated sensitivity(d') and criterion(c) for each subject and block using the following standard formulas where H is the normalized percent of accurate matches and FA is the normalized rate of false alarms on no-match trials:

$$d' = H - FA \tag{4.1}$$

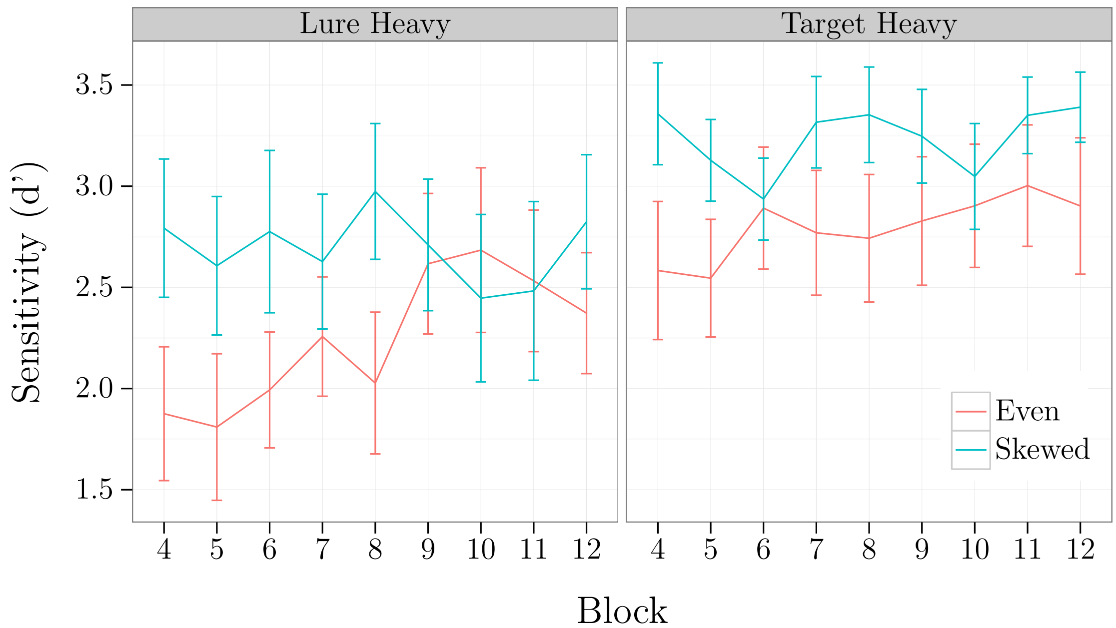
$$c = -(H + FA)/2 \tag{4.2}$$

An ANOVA of the full dataset revealed a main effect of block on sensitivity (d'), $F(11{,}770)$=9.60 $p<.001$, with significant linear ($p<.001$) and quadratic ($p<.001$) trends. The full ANOVA table is located in the appendix (table C.2). A subsequent analysis performed on the truncated (blocks 4-12) dataset revealed no significant main effect of block (p>.2). The ANOVA table for this truncated analysis can also be found in the appendix (table C.3) The lack of a significant trend for the truncated dataset suggests that the trends found in the full analysis were caused by practice effects during the initial 3 blocks. Thus the remaining analyses likely reflect steady state performance (e.g. differences between conditions are not due to learning or practice effects.)

A similar process was performed for the criterion values (c) and again found no significant linear effect of block in the truncated dataset. The statistics cited in the remainder of this section will be limited to the truncated datasets, reflecting steady state performance. ANOVA tables for both the full and truncated datasets can be found in the appendix (tables C.4 and C.5).

T:L ratio had a significant effect on sensitivity(d'), $F(1{,}70)=5.04$ $p<.03$, $\eta_p^2=.07$. Target heavy subjects (M=3.02, SD=1.09) outperformed lure heavy subjects (M=2.51, SD=1.41). Stimulus distribution had a marginal effect on sensitivity $F(1{,}70)=3.46$ $p<.07$, $\eta_p^2=.047$.

T:L ratio also had a strong effect on criterion (c), $F(1{,}70) = 23.16$, $p<.001$, $\eta_p^2=.25$. Subjects in the Lure Heavy condition (M=.54, SD=.58) were more biased toward answering no-match than were subjects in the target heavy sequences (M=.17, SD=.30). Sequence skew had no effect on criterion. Overall, the signal detection analysis suggested that subjects in the lure heavy groups developed a significant bias toward answering no-match. This was unexpected, but makes sense since 56 out of 64 trials per block are no-match trials in these conditions.



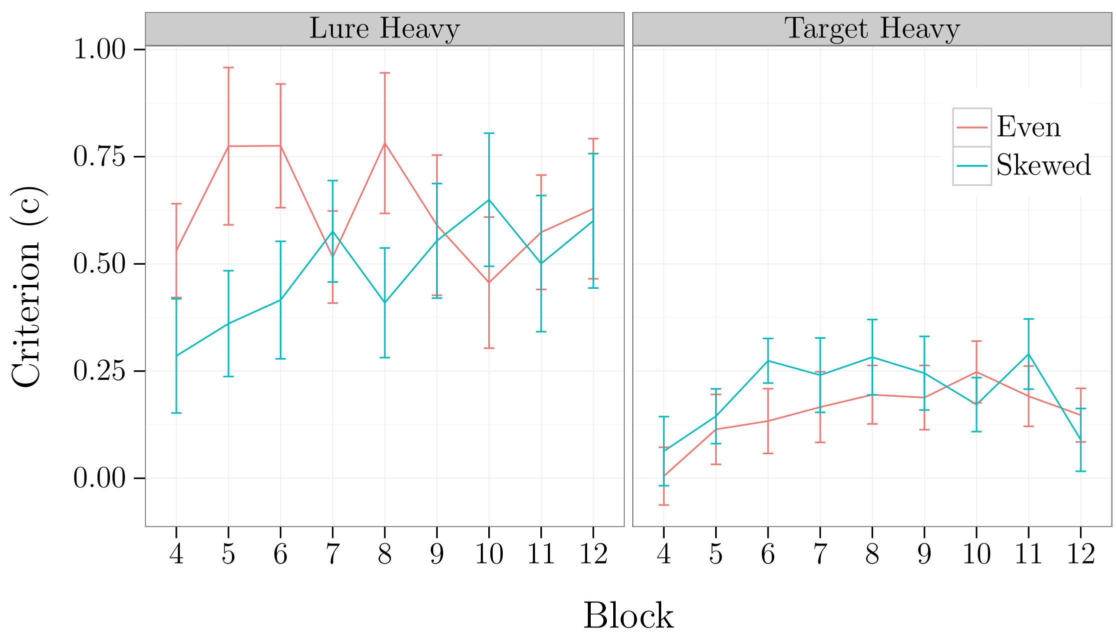


Figure 4.2: Top: Sensitivity (d') by block. d' measures the difference in units of standard deviation between correct responses on match trials and false alarms on non-match trials. Bottom: Criterion (c) is a measure of response bias. Values closer to 1 denote a bias to respond "no", Values closer to -1 denote a bias to respond "yes". 0 represents no bias. Error bars represent standard error.

Although the signal detection analyses revealed main effects of the independent variables, they are not sufficient to determine strategic differences caused by sequence factors. Figure 4.3, a plot of accuracy by serial position, suggests that the signal detection main effects were likely related to strategy differences. Lag is the number of trials since the current letter was last seen. Thus lag position 3 are match trials, and positions 2 and 4 are lure trials. The remainder of the work will discuss position 2 (n-1) and position 4 (n+1) lures separately. I will refer to them as n-1 lures and n+1 lures, respectively. The drop in accuracy at serial position 3 (matches) in the lure heavy plot is likely caused by the significant bias toward answering no-match. This bias also resulted in improved performance on lure trials when compared to the target heavy sequences.

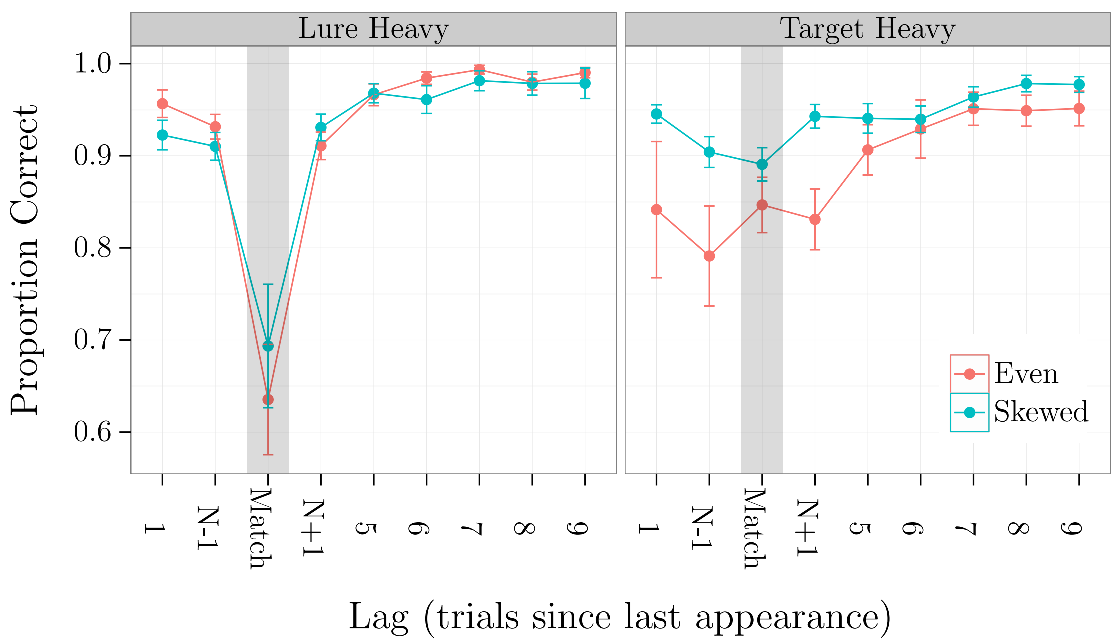


Figure 4.3: Accuracy by lag. Data presented is accuracy rates for each sequence type by lag, the number of intervening trials since the current letter was last seen. Position 3 is the match position, and positions 2 and 4 are lure trials. The left side compares lure heavy (even-8 and skewed-8) sequences. The right side compares target heavy (even-24 and skewed-24) sequences. Error bars represent standard error.

4.2.2 Response Time

Response times are often seen as a more informative measure of control and strategy differences than accuracy. Different strategies may result in similar response accuracy, but since they often rely on different cognitive processes, response time differences tend to emerge. Proactive control should lead to faster response times on match trials than lure trials because it involves activating the target item in anticipation of stimulus onset. Lure trials should be slower because of conflict caused by the increased activation of the lured letter due to rehearsal. Reactive control, by contrast, does not cause differences in response time between match and lure trials. Reactive control does not involve added rehearsal/activation of the recent letters, which is thought to be responsible for differences in response times (Szmalec et al., 2011).

Figure 4.4 displays response time by lag position. Target heavy sequences exhibited the proactive pattern described above. Match trials (labeled 'Match' in the plot) are faster than lure trials (n-1 and n+1). By contrast, lure heavy sequences do not exhibit this pattern. These data do not support the hypothesis of proactive control for lure heavy sequences and reactive control for target heavy sequences. On the contrary it seems that proactive control was used for target heavy and not lure heavy sequences.

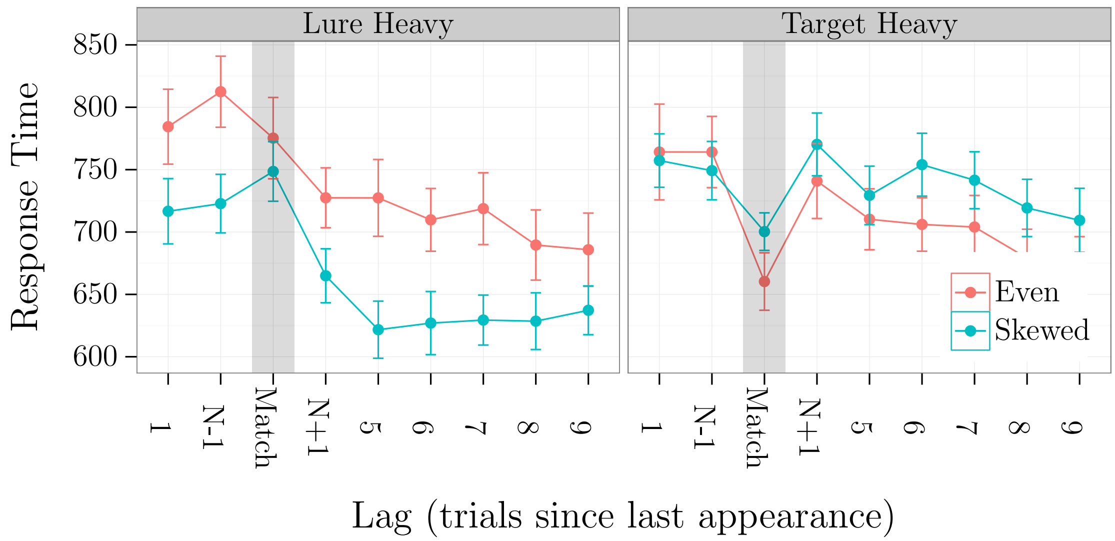


Figure 4.4: Response time by lag position. The shaded data point is position 3, which corresponds to match trials. Error bars represent standard error.

In order to quantify the difference between match and lure trials, I subtracted match response times from n-1 and n+1 lure response times, respectively. A similar practice is common in studies examining control issues using the flanker and stroop task (Yeung, Bogacz, et al., 2004; Yeung, Ralph, & Nieuwenhuis, 2007; Botvinick et al., 2001) where the mean response time on congruent trials is subtracted from the mean response time on incongruent trials Task switching studies also use a difference approach in calculating the switch cost (the difference between response times on trials where the task switches vs repeats) (Altmann & Gray, 2008).

Figure 4.5 displays these data. The left panel shows the difference between n-1 lures (those last seen 2 trials in the past) and matches. The right panel displays the difference between n+1 lures (last seen 4 trials previously) and matches. Negative values occur when lure response times are faster than match response times.

The distinction between n-1 and n+1 lures is extremely important. Assuming a Baddeley style rehearsal process, n-1 lures are in the current rehearsal loop, but n+1 lures are no longer being rehearsed. Thus although excessive rehearsal can cause increases in response times on n-1 lures, this is less likely to affect performance on

n+1 lures, which are no longer being rehearsed. Increases in n+1 response times, by contrast, may be a residual effect of proactive control applied to the target item on the previous trial.

Trend analysis revealed a significant linear trend of response time by block, but there was no interaction related to any of the factors (T:L ratio or Stimulus distribution). Thus subjects continued to respond more quickly as time passed, but this increase did not seem to relate to strategic or environmental differences across conditions. In order to simplify the response time analyses, I collapsed across block and performed an ANOVA on the difference between lure and match response times. The design of the ANOVA was 2 T:Lure ratio (Target Heavy/Lure Heavy) x 2 Stimulus Distribution (Even/Skewed) x 2 Lure position (n-1/n+1). The full ANOVA table is in the appendix (table C.7).

The results of the analysis are best understood by considering figure 4.5. The ANOVA revealed main effects of T:L ratio $F(1{,}70) = 22.67$, $p<.001$, η_p^2=.24 and Lure Position $F(1{,}70) = 11.85$,p=.001, η_p^2=.14. Critically, the analyses also found significant interactions. T:L ratio interacted with Lure Position, $F(1{,}70) = 32.26$,$p<.001$, $\eta_p^2 = .31$. Specifically, T:L ratio affected the rt difference more strongly when considering n+1 lures than n-1 lures. By contrast, stimulus distribution interacted with lure position in the opposite direction, affecting n-1 lures more strongly than n+1 lures, $F(1{,}70) = 14.23$, $p<.001$, $\eta_p^2 = .16$.

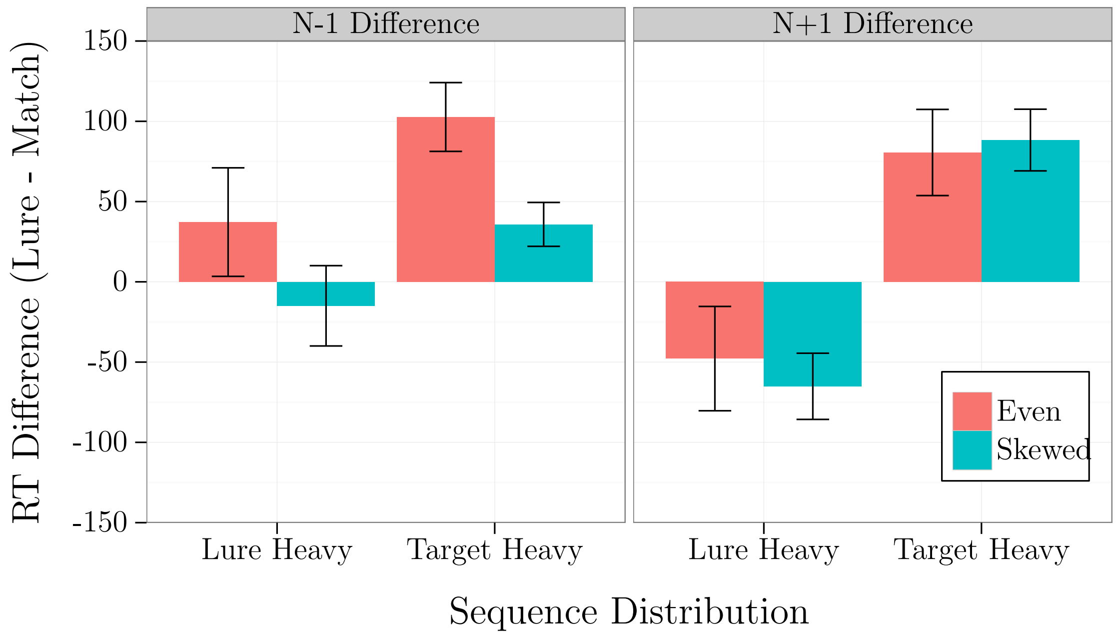


Figure 4.5: Plots of the response time difference between lure and match trials. Left Panel: differences between n-1 lures and matches. Right panel: differences between n+1 lures and matches. Error bars represent standard errors.

The large effect sizes found for the interactions between lure position, stimulus distribution and T:L ratio provide evidence of a dissociation between control mode and rehearsal strategy. The n+1 results suggest a difference in the control mode between target heavy and lure heavy sequences. Contrary to the original hypothesis, these data support the d' accuracy result suggesting that lure heavy subjects primarily used reactive control while target heavy subjects used proactive control. Because n+1 lures are not in the current rehearsal window, the increase in activation (and resulting conflict) is most likely a result of the activation applied on the previous trial (when it was the match trial). In other words, the more proactive control used (increasing activation of the 3rd back item), the more difficult it was to reject that item when it was presented on the subsequent trial (as an n+1 lure). The data depicted in the right column of figure 4.5 show that this occurred only for target heavy sequences, suggesting that subjects presented with lure heavy sequences did not use proactive control to increase the activation of the target item.

The n-1 results, on the other hand, suggest a difference in the amount of rehearsal used. n-1 lure trials are in the current rehearsal window, so differences in response times should be correlated with the amount of rehearsal these items receive. The data depicted in the left panel of figure 4.5 suggest that subjects rehearsed more when presented with even sequences.

Since the distribution of stimuli effected the amount of rehearsal used, it is possible that subjects adjusted their use of rehearsal as local lumpiness changed. Refer to chapter 3 for my discussion of how lumpiness varied within sequences. I fitted a linear model of the response time for n-1 lure trials as a function of local lumpiness. There was a significant interaction between stimulus distribution and lumpiness, $F(1,367) = 15.05$, $p<.001$. The full ANOVA table derived from the linear model can be found in the appendix (table C.8).

This interaction can be best understood by considering figure 4.6. Response times for target-heavy skewed sequences decrease as a function of lumpiness, while response times for even sequences do not vary with lumpiness. This interaction reveals an important piece of evidence relating to control. The decrease in response time with increasing lumpiness for target heavy skewed sequences is consistent with the explanation that these subjects adjusted the amount of rehearsal as sequences became more lumpy. Less rehearsal with increasing lumpiness would have reduced the additional activation provided to lure items, leading to faster response times.

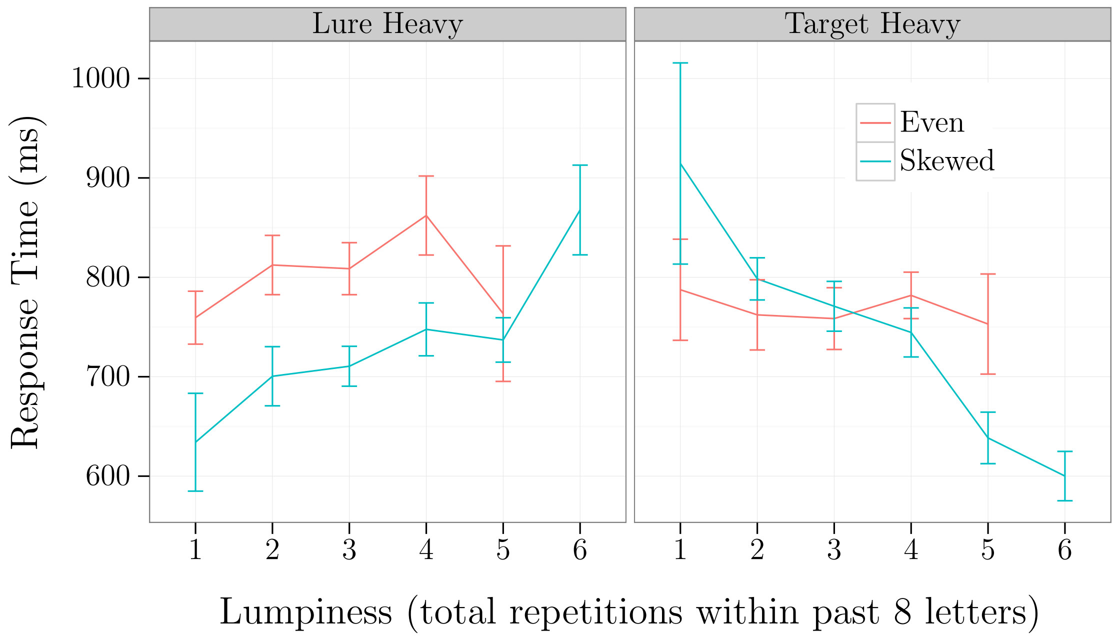


Figure 4.6: Plots of response time by local lumpiness for n-1 lure, n+1 lure, and match trials. Lumpiness (x-axis) represents the number of stimulus repetitions found within the previous 8 trials. Error bars represent standard errors.

4.3 Discussion

The results from this experiment provided evidence of control variation in the n-back task. Several of the results supported the hypothesis of increased proactive control for even sequences when compared to skewed sequences. The results did not, however, support the hypothesis of increased proactive control for lure heavy sequences. In fact, the opposite effect seemed to occur, with target heavy conditions exhibiting signs of proactive control and lure heavy conditions showing markers of a reactive, activation based strategy. While it is tempting to make conclusions regarding control mode based on these data, variation in strategy use (rolling or static) must also be considered, as well as factors not controlled for, such as the existence of trial by trial feedback. Experiments 2 and 3 build upon the foundation of experiment 1 by examining the effect of different strategies (e.g. rolling vs static) on performance.

4.3.1 Stimulus Distribution

The results provide strong support for the hypothesis that stimulus distribution affects control mode. Subjects in skewed conditions reacted more strongly to differing levels of lumpiness than those in even conditions. By definition, a reactive strategy will be more responsive to changes in the task environment than a proactive one. Thus the finding of an effect of lumpiness on response times only for skewed sequences supports the hypothesis that skewed sequences bias people toward a reactive control mode. These subjects seemed to modify the amount of rehearsal they used based on the state of the environment. In fact, the lack of an effect of lumpiness for even sequences supports the hypothesis of increased proactive control for these sequences. Proactive control does not rely on the activation provided by the repetition of common letters because rehearsal keeps the last 3 items highly active regardless of how often they have been repeated.

The analysis of the difference in response times between match and lure trials provided additional evidence as to how stimulus distribution affected control mechanisms. It is unclear if these results reflect a difference in control mode (proactive or reactive) or control strategy (static or rolling). Stimulus distribution affected the difference between lure and match RTs only for n-1 lures.

Further, local lumpiness (defined as the number of repetitions among the previous 8 trials) co-varied with n-1 response times for skewed sequences only. These results reveal a strong effect of the task environment on control processes. Subjects adjusted their use of control (rehearsal) to the immediate environment (e.g. the past 8 trials), but only for skewed sequences.

Thus the effect of stimulus distribution caused a change in the amount of rehearsal used, but since distribution did not affect n+1 RT differences, there is no evidence that this extra rehearsal for even subjects translated into effective proactive control. Indeed, although response times on match trials were faster for even than skewed sequences, even sequences resulted in lower accuracy for match trials. Based on these results we can conclude only that skewed sequences allowed subjects to reduce the amount of rehearsal performed without impairing their ability to recognize matches. Further studies are necessary to better understand how distribution affects the choice of strategies.

4.3.2 T:L Ratio

In contrast to the effect of stimulus distribution, the results provided no support for the T:L ratio hypothesis. Instead of leading to a bias toward proactive control, lure heavy sequences seemed to bias subjects toward a reactive mode. This was surprising given the substantial literature demonstrating a strong effect of the ratio of incongruent to congruent trials in many tasks (Braver, Gray, & Burgess, 2007; Botvinick et al., 2001; Yeung, Bogacz, et al., 2004; Yeung, Ralph, & Nieuwenhuis, 2007; Bailey et al., 2010). People generally opt for a proactive strategy when the likelihood of conflict rises. The data from the current experiment do not exhibit this common feature.

One explanation is that people adjusted to the reward structure of the environment. Each lure heavy sequence featured 56 “no-match” and only 8 “match” trials. Thus an optimal strategy might be one that simply maximizes the chance of correctly rejecting “no match” trials. In fact, subjects would have responded correctly on 87.5% of trials by simply responding no to all trials. The trial by trial feedback provided could have reinforced this reward based strategy by making it clear that very few match trials actually occurred. Even though half of the trials (24 lures and 8 targets) featured similar levels of activation/familiarity, the feedback provided may have made the fact that only a small minority were actually matches more salient than had there been no feedback. The signal detection analysis of the lure heavy sequences suggested that subjects were in fact sensitive to the increased likelihood of “no match” trials. The criterion value (figure 4.2 bottom panel) was significantly higher on lure heavy sequences than target heavy sequences, suggesting a strong bias to respond no.

Results from several studies in the literature provide support for the view that n-back subjects are sensitive to the ratio of matches to no-matches. Much like the lure heavy conditions, which contained 12.5% match trials, Harbison et al., 2011 presented very few match trials per block (20%). Although Harbison did not report exact accuracy results in his conference paper, estimates derived from visually inspecting his plot of accuracy by trial type (60% correct for match trials, ˜85% correct for no-match trials) yielded a criterion value of .41, largely in line with the results from the lure heavy sequences in the current study. Harbison's subjects were even less accurate on matches (60%) than in the current study (67%). Szmalec 2011, on the other hand, presented many more matches per block (33.3%) and results from his experiment 1 (87% correct for match trials, 93% correct for no-match trials) yielded a criterion of .17, in line with the target heavy results from the current experiment.

Both the Harbison and Szmalec studies would be considered target heavy (despite the small number of match trials in the Harbison study, there were even fewer lures), so the difference in criterion was not a function of the T:L ratio, but rather the total ratio of matches to non-matches. I suspect the same effect explains the bias found in the current experiment. Although the results of the current experiment provide no support for the hypothesis of increased proactive control for lure heavy sequences, it is possible that an unexpected factor (ratio of matches to non-matches) resulted in an even stronger bias in the opposite direction. Experiment 2 will test this possibility by removing the trial by trial feedback, obscuring the true number of matches/non-matches.

4.3.3 General Discussion

The current experiment succeeded in providing evidence for control variation in response to specific manipulations of the stimulus sequences in the n-back task. In addition, the results suggested avenues for future research. The results provided support for the hypothesis of increased proactive control for even sequences. Lure heavy sequences, on the other hand, did not provide the expected bias toward proactive control. However, unexpected factors such as match/no-match ratio and the existence of feedback may have corrupted the ability to study the effect of T:L ratio.

Strategic variation was not investigated per se in the current experiment, but a 10,000 ft view of the results do lead to an explanation that involves strategy choice. The response time data suggested that subjects in even conditions used more rehearsal than those in skewed conditions. Subjects seem to react to changes in sequence lumpiness, but only when presented with the skewed sequences. It is possible that skewed and even sequences lead people to use different types of strategies (e.g. rolling vs static).

The remaining experiments in the current work will focus on how strategic differences interact with the stimulus distribution and T:L ratio. Experiment 2 holds stimulus distribution constant (even sequences) in order to measure the interaction of strategy selection, trial by trial feedback, and T:L ratio. Conversely, experiment 3 holds T:L ratio constant (target heavy sequences) in order to measure the interaction of strategy selection and stimulus distribution.

Chapter 5
Experiment 2

5.1 Purpose

The results from experiment 1 supported the general hypothesis that sequence factors affect the use of proactive control, but they did not support the specific prediction that lure heavy sequences bias people toward a proactive control mode. These sequences instead seemed to lead to a reactive strategy. Like previous research that used a small proportion of match trials, subjects responded more accurately on non-match trials, but struggled to correctly identify matches (33% errors). Additionally, analysis of response times showed that RT varied as a function of position lag, with no additional decrease in RT for match trials, which did occur for target heavy sequences.

Although the results suggest a reactive strategy, it remains unclear if other factors such as the existence of trial by trial feedback could have contributed to strategy selection. Trial by trial feedback may have made the reward structure of the environment more salient that it otherwise would have been. This explanation fits within the bounded optimal control strategy proposed by Lewis et al (Lewis, Shvartsman, & Singh, 2013), in which eye movement strategies were selected based on the reward structure of the environment. Subjects could have learned very quickly that the vast majority of recently seen letters are in fact not matches. Proactive rehearsal would have increased the activation of these common lure trials, making them more difficult to reject, with the only benefit being an enhanced ability to recognize the rare match trials. In experiment 2 I removed the feedback to test the inverse of Lewis's bounded optimal control theory, namely that obscuring the nature of the reward structure will reveal biases inherent to cognitve control processes.

A second possible confound not addressed in experiment 1 was the effect of rehearsal strategy selection. Some subjects likely used the static strategy while others likely used rolling rehearsal. In fact, it is not clear that subjects "chose" a strategy at all, as people are often unaware of cognitive control mechanisms.

Differences in the dynamics of rehearsal strategies could have created variation in the response time measures. It is possible, for instance, that subjects were more likely to use the rolling rehearsal on even sequences. This would explain the evidence of extra rehearsals performed by even vs skewed subjects. Experiment 2 controlled (as best as possible) the use of a specific strategy. Because it is the easiest strategy to describe and is likely used by most people, I trained half of the subjects to use rolling rehearsal (and allowed the other half to perform the task without suggestion).

5.2 Methods

5.2.1 Participants

72 subjects were recruited from the RPI campus and were awarded course credit in exchange for their participation. The design of the experiment called for 15 subjects per condition for a total of 60. 12 additional subjects were added at the end of the experiment to replace experimental sessions lost due to malfunctioning equipment.

5.2.2 Materials

Materials were identical to those used in experiment 1 with the following change. All feedback (colored boxes and block summary feedback) was removed so that subjects could not use feedback to modify strategies.

5.2.3 Design

The design of this experiment was 2 (strategy) x 2 (T:L ratio) x 12 (block) with strategy and T:L ratio manipulated between subjects. Following a 3 mini-block training procedure (discussed below), each subject completed 12 experimental blocks. Subjects were randomly assigned to one of 4 between-subjects conditions. Half of the subjects were given explicit instructions about how to use the rolling rehearsal strategy and were asked during practice to rehearse out-loud so that the researcher could verify that subjects were in fact using the strategy correctly. The other half were given no instruction regarding strategies and were instead simply provided with the rules of the task (e.g. determine if the current letter is the same as the 3rd previous letter ...)

Each group of subjects (trained or control) was further divided into two groups. One group was presented with lure heavy sequences, and the other group target heavy sequences. The conditions were Rolling-24, Rolling-8, Control-24, and Control-8, reflecting whether or not they were trained to use the rolling strategy and the number of match trials per block. The replacement of subjects due to computer malfunctions and experimenter error in assigning replacement conditions resulted in a slightly unbalanced design. During data analysis it was discovered that one subject stopped responding during the 12th and final block, so that subject was also removed. The result was a total of 59 subjects (subject counts by experimental condition are displayed in table 5.1).

Table 5.1: Sequence Conditions for Experiment 2

Condition	Targets	Lures	Distribution	T:L Ratio	Subjects
Rolling-8	8	24	Even	1:3	13
Rolling-24	24	8	Even	3:1	15
No-training-8	8	24	Even	1:3	16
No-training-24	24	8	Even	3:1	15

In order to allow for an exact comparison with experiment 1, experiment 2 used the identical even sequences from experiment 1. I chose to use the even sequences because the results from experiment 1 suggested that these sequences led to the use of proactive control strategies. Since one goal of experiment 2 was to examine the rolling rehearsal strategy, it seemed best to limit the investigation to these even sequences.

5.2.4 Procedure

Each experimental session began with a short introduction to the task. Researchers presented each subject with a short slide-show explaining the basics of the n-back task. A copy of these slides are included in the appendix. The presentation explained that subjects needed to determine whether each presented letter was the same or different as the third previous letter. The researcher guided the subject through a short sample sequence, pointing out which trials were matches and which were not matches.

Subjects in the rolling condition completed a 3 mini-block training procedure. Researches urged subjects to focus on performing the instructed strategy regardless of resulting performance and even if they felt there was a better way to do it. Researchers then gave detailed instruction to subjects on the use of the rolling strategy. Presentation slides for this instruction are also included in the appendix. Subjects in the no-training group received only rudimentary instruction consisting of a basic description of the task (e.g. determine if the current letter is the same or different than the third previous letter).

All subjects (including the control group) performed three 32-trial practice blocks [2]. The experimenter remained in the experimental room during these training blocks. The first practice block was self-paced with feedback. Subjects could take as much time as necessary to make responses. The lack of time pressure gave the experimenter a chance to reinforce the trained strategy and confirm that the subject was in fact using the strategy correctly. Subjects in the rolling groups were asked to report the current list of the previous three items in the correct order out-loud after each letter appeared so that the researcher could confirm that they were using the correct strategy (e.g. BCD if the last 4 were A,B,C,D). As in experiment 1, a colored box appeared after each response. A red box indicated an error and a green box followed a correct response.

The second practice block also provided feedback via colored boxes but was not self-paced. Instead, the second practice block followed the timing of the experiment (letters appeared with a 3000 ms SOA). This block provided subjects with the opportunity to get accustomed to the time pressure of the study while still practicing the instructed strategy. In common with practice block 2, the third practice block also followed the standard timing of the experiment, but in contrast to practice

[2] Due to an oversight, the experiment presented all subjects with the same 32 trial practice blocks. These practice sequences were taken from a target heavy even sequences. Thus all subjects practiced with a target heavy sequence. This mistake had the potential to effect the learning process of the subjects in lure heavy conditions, as they practiced with a different type of sequence than the sequences they would experience in the experimental session. This mistake should be moderated by the elimination of the first 3 experimental blocks from the analysis as will be discussed in the results section.

blocks 1 and 2 did not provide feedback after each response. This practice block gave subjects a final chance to practice the strategy and prepare for the experimental blocks, which except for the difference in the number of trials are identical in design (no feedback, 3000 ms SOA).

At the end of the 3rd practice block subjects in the rolling groups were asked to rate their ability to perform the rolling strategy on a scale of 1 to 7. They were then asked if they would like to perform an additional practice block. If chosen, the additional practice block was identical in design to the 3rd practice block. 12 subjects chose to complete a 4th practice block. These included 4 Rolling-24, 3 Rolling-8, 3 no-training-24, and 2 no-training-8 subjects.

Subjects in the control (no-training) groups performed the same 3 practice blocks as the rolling groups but were not asked to rehearse out loud. Researchers only intervened if subjects made frequent mistakes in the first practice block. Intervention consisted only of asking if the subject understood that they were to determine if the current item matched the 3rd previous item.

The experimental phase of the session was divided into 12 blocks, with each block including one 64-trial sequence. As in experiments 1, all sequences sampled from the following 8 phonologically distinct letters: B,F,H,J,M,R,X,Z. In fact the sequences used were identical to the 24 even sequences (12 target heavy and 12 lure heavy) used in experiment 1. The use of identical sequences made possible a comparison between analogous conditions in experiments 1 and 2. The no-training conditions in experiment 2 matched the even sequences from experiment 1 with the only differences being the existence (or lack thereof) of trial by trial feedback, and a difference in the number of practice blocks. The difference in practice should be moderated by truncating the first 3 experimental blocks from data analysis (as was done for all experiments).

At the end of the session subjects were asked several questions based on the experimental condition. Subjects in the rolling conditions were asked to rate their success at performing the instructed strategy. They were also asked if they found themselves using a different strategy and to describe the alternate strategy. After describing an alternate strategy, the researcher determined if the described strategy matched either the static or rolling rehearsal strategies or a hybrid version. In order to do this, the experimenter presented the subject with a hypothetical subsequence and asked the subject what they would do next (e.g. “if the last three letters were R, D, F, and the new letter is a C, what would you do next”). If the subject indicated that her new memory list would be D,F,C, the experimenter coded the strategy as rolling. If the subject replied C,D,F, the experimenter will coded the strategy as static. Using a similar procedure, researchers asked no-training subjects to describe the memory strategy used in their own words. Again using the procedure described above, the experimenter coded the strategy as rolling, static, or hybrid.

5.3 Results

The results of experiment 2 will be presented in two sections. The first analyses performance (accuracy and RT) between the 4 groups of between-subjects conditions. The second section further analyzes the control condition by grouping these subjects based on their responses during the post experiment interviews.

5.3.1 Basic Analysis

5.3.1.1 Response Accuracy

As in experiment 1, trend analyses were performed for the full (12 block) and truncated (final 9 blocks). Significant linear and quadratic trends found in the full dataset for d’ were again eliminated by truncating the first 3 blocks. A significant effect on criterion values remained and is discussed below. The full ANOVA tables for each analysis can be found in Appendix C (tables C.2 -full, C.3 - truncated).

The truncated ANOVA showed no significant effects of T:L ratio $F(1{,}56)$ = .02 p>.8 and a marginal effect of strategy training $F(1{,}56)$ = 3.31, p=.07, η_p^2=.06. on sensitivity (d’). As table C.1 shows, trained subjects were slightly more accurate than non-trained subjects. The lack of a main effect of T:L ratio fails to replicate the effect found in experiment 1. Figure 5.1 shows that non-trained subjects in experiment 2 did not adopt the same biased reactive strategy of their experiment 1 counterparts. The conditions plotted are identical with respect to all experimental design parameters except for the existence of trial by trial feedback in experiment 1.

Focusing on the lure heavy sequences (left side), experiment 1 subjects were more accurate on lure and distractor trials and less accurate on match trials than subjects from experiment 2. These data support the hypothesis that the strategy employed by subjects in the lure heavy conditions of experiment 1 adopted a reactive strategy in response to the reward structure of the environment. That reward structure was not made explicit in experiment 2, and these accuracy by lag data show no differences between target heavy and lure heavy sequences.

Further, in Experiment 2, neither T:L ratio $F(1{,}56) = .57$ $p>.4$ nor strategy training $F(1{,}56) = 1.32$ $p>.25$ affected criterion values. Criterion values did, however, rise over time $F(8{,}448)=2.197$, $p<.03$, $\eta_p^2=.04$. There was a significant linear trend found in the 3-way interaction of T:L Ratio, Strategy and Block $F(1{,}448)=7.25$, $p<.01$. As figure 5.2 shows, this effect is likely due to the increase in criterion value for the rolling, lure heavy group during the final 2 blocks of the experiment. Although this trend is statistically significant, it is unclear how it would relate to the hypotheses of the current work.

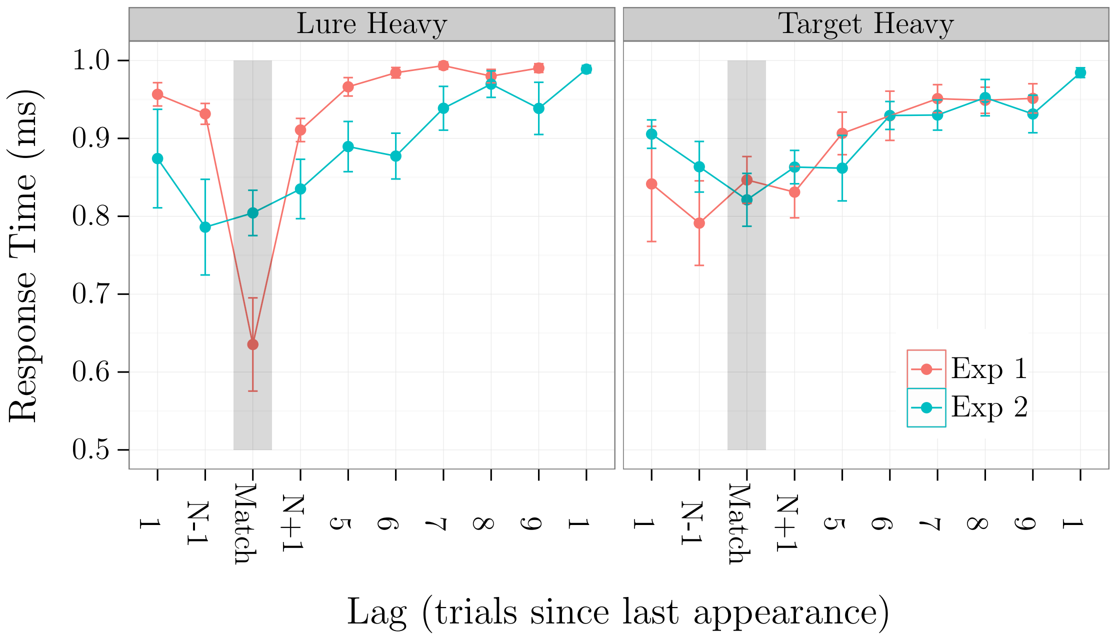


Figure 5.1: Comparison of accuracy by lag for experiments 1 and 2. Data from the even sequences of experiment 1 (red) and non-trained subjects of experiment 2 (blue). These conditions were identical except for the existence of trial by trial feedback. Error bars represent standard error.

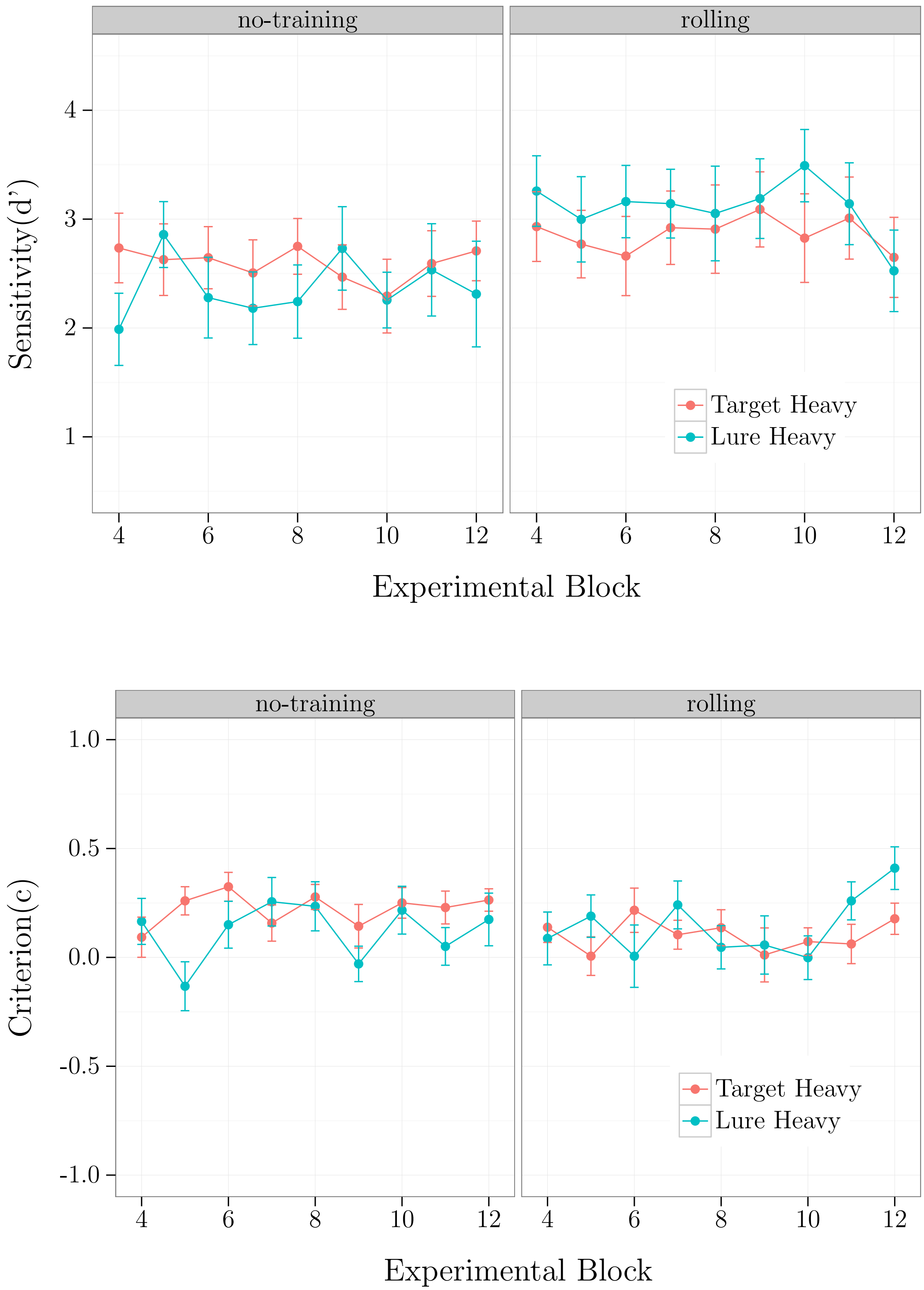


Figure 5.2: Plots of sensitivity (top) and criterion (bottom) by block. Sensitivity is shown in units of d'. Criterion values range from 1 to -1, with higher values relating to a bias to respond no. Error bars represent standard error.

5.3.1.2 Response Time

Response time data also provided evidence related to the effect of feedback in experiment 1. The response time by lag plots of figure 5.5 suggest a proactive component for the lure heavy sequences that was not found in experiment 1. Here, match trials are faster than n-1 and n+1 lure trials, suggesting subjects were in fact preparing for their appearance ahead of time. This contrasts with the data for the identical sequences of experiment 1 in which match trials were slightly faster than n-1 lure trials and slower than n+1 lure trials. The evidence for proactive control among lure heavy sequences in experiment 2 applies regardless of strategy training condition, so the difference is likely due to the lack of trial by trial feedback. It must be noted that although there was some proactive control used by lure heavy subjects in experiment 2, target heavy subjects still exhibited more evidence of proactive control. Thus the feedback provided in experiment 1 likely made the reward stucture of the environment immediately explicit, creating a bias to respond no. Subjects in experiment 2 did not exhibit this bias, but were still less proactive than their target heavy counterparts. The effects can be clearly seen in figure 5.3. Figure 5.3 is a comparison of complementary conditions in experiment 1 and experiment 2.

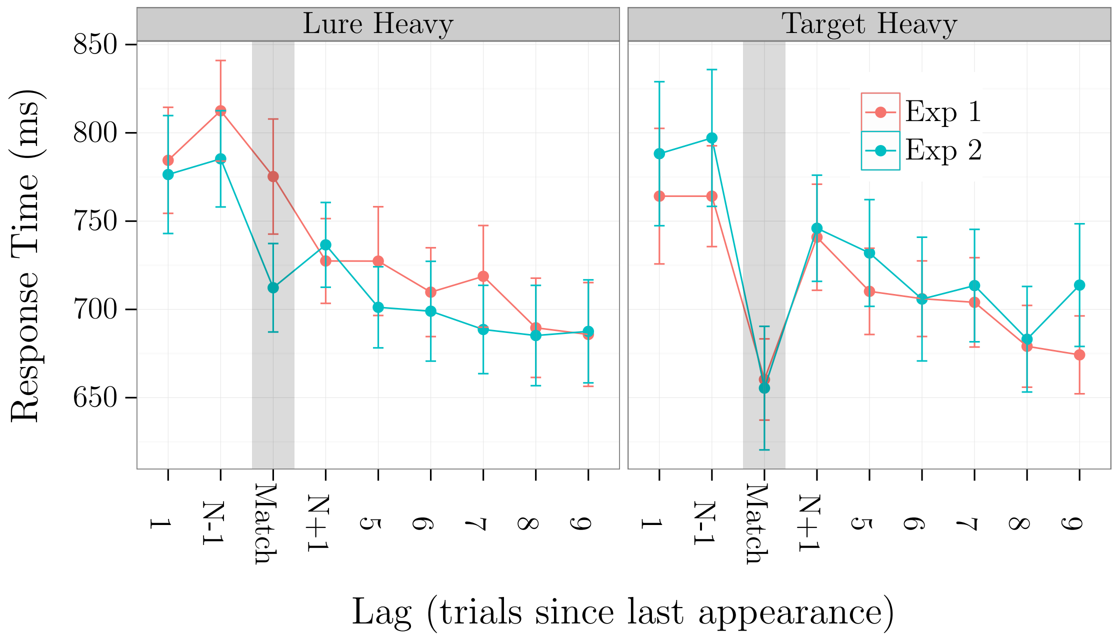


Figure 5.3: Comparison of RT by lag for experiments 1 and 2. Data is for the even sequences from experiment 1 and the non-trained subjects from experiment 2. Error bars represent standard error.

As in experiment 1, the difference in response time between matches and lures helped to determine control mode differences. Figure 5.4 displays these results. A 2 (Training condition) x 2 (T:L ratio) x 2 (lure position) ANOVA (table D.7) revealed that target heavy conditions led to larger differences than lure heavy sequences, $F(1{,}56) = 15.00$, $p{<}.01$, $\eta_p^2{=}.22$. This large effect size replicates the finding from experiment 1, which had a similar effect size ($\eta_p^2{=}.24$). Again, despite evidence of some proactive control for lure heavy sequences, these proactive influences were much stronger for target heavy sequences, replicating the results of experiment 1. Interestingly, The ANOVA failed to find a significant interaction between T:L ratio and lure position. This comparison was significant with a large effect size ($\eta_p^2{=}.14$) in experiment 1. As can be inferred from figure 5.3, the difference likely stems from the drop in RT for match trials for lure heavy subjects in experiment 2. Again, this difference provides evidence of the use of proactive control processes in by lure heavy subjects in experiment 2.

The analysis of response time differences also shed light on the effectiveness of the rolling rehearsal strategy. The ANOVA revealed that the difference between response times on matches and n-1 lures was larger than the difference between matches and n+1 lures, $F(1,56) = 68.13$, $p<.001$, $\eta_p^2=.54$. Training condition (rolling or not trained) interacted with lure position, $F(1,56) = 5.11$, $p<.03$, $\eta_p^2=.08$. Figure 5.5 shows that the difference between n-1 lure and match trials was due to slower response times on n-1 trials, not faster responses on match trials as was the case in experiment 1. This provides evidence that the rolling strategy is a particularly ineffective rehearsal strategy which causes unnecessary conflict on lure trials without improving performance on match trials.

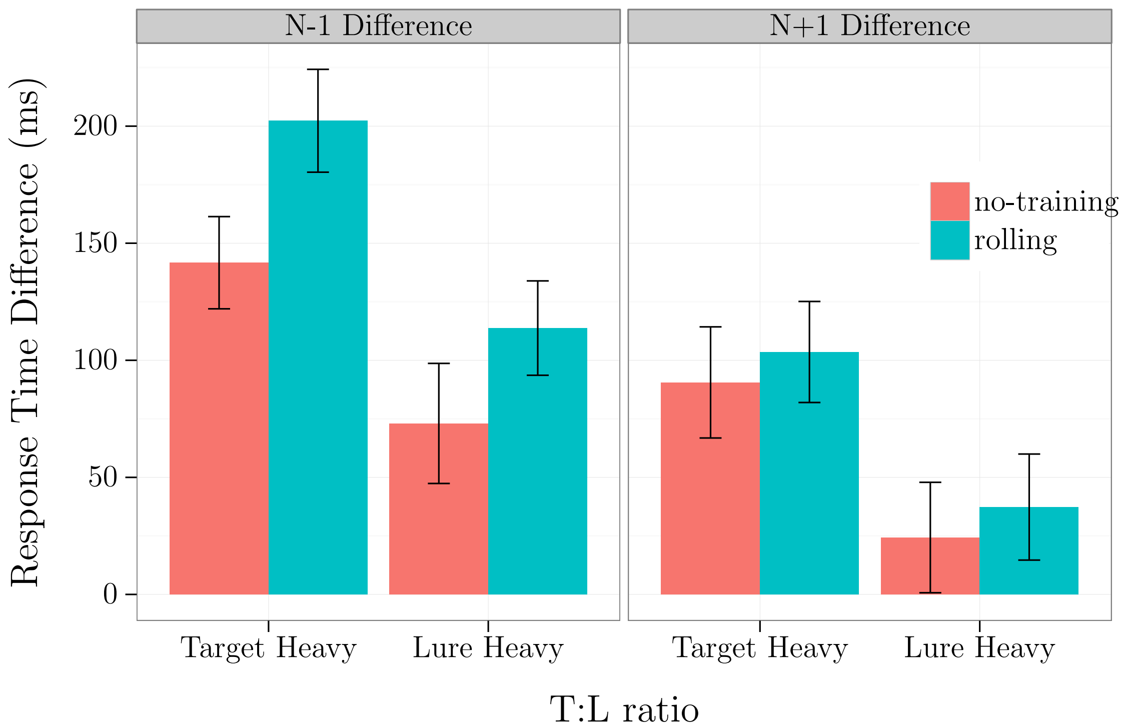


Figure 5.4: Response Time differences between target trials and lure trials, separated by lure type. Right: Response Time differences between target and n-l (2-back) lure trials. Left: Response Time differences between target trials and n+1 (4-back) lure trials. Error bars represent standard error.

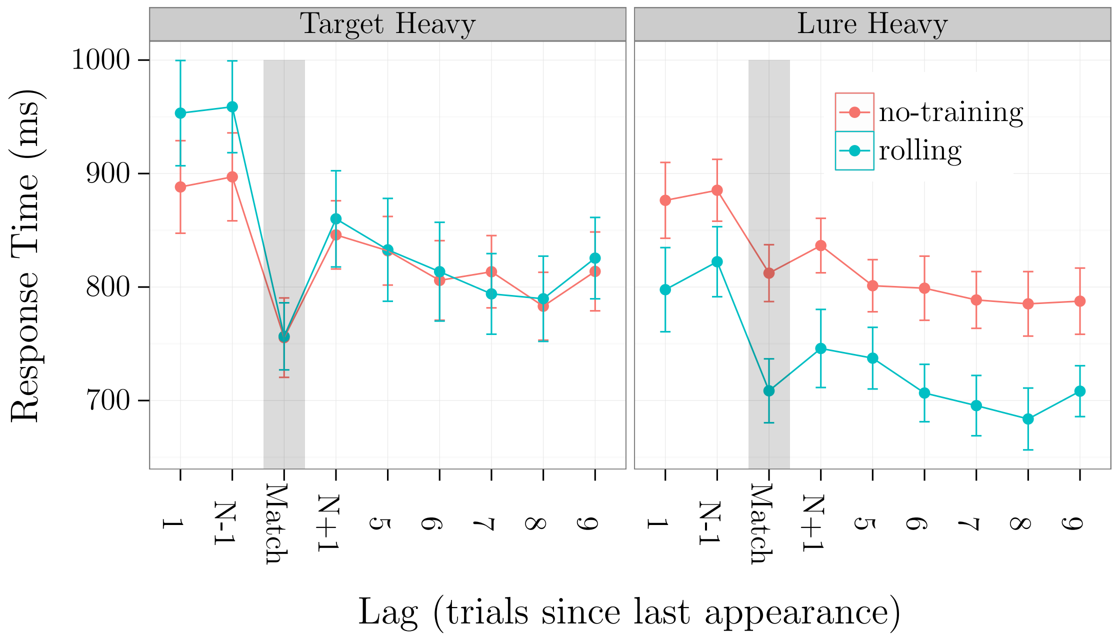


Figure 5.5: Response times by lag position for all subjects. Target heavy data is plotted on the left and lure heavy data on the right.

In experiment 1, the difference between skewed and even sequences on n-1 lures related to the amount of rehearsal used. Skewed sequences showed a sensitivity to the current state of the sequence (lumpiness), while even conditions showed no relationship between n-1 lure response times and lumpiness. Experiment 2 resulted in a significant effect of training condition on n-1 lure RT differences, but since these are all even sequences, It is unlikely that subjects adjusted to lumpiness, as was the case for the skewed sequences in experiment 1. An analysis of covariance showed no main effect of lumpiness (p=.27) or interaction between training and lumpiness (p=.47). The lack of significant effects of lumpiness replicates the finding in experiment 1 with respect to even sequences. Since lumpiness did not effect n-1 response times, it is likely that strategic variation, and not the amount of rehearsal, is responsible for the significant interaction between training condition and n-1 response time difference.

5.3.2 Strategy Differences

I included the control (non-trained) groups in order to compare the use of a specific strategy (rolling) to performance without a trained strategy. The responses to the post-experiment interviews, however, made it clear that non-trained subjects were a heterogeneous group. Although a majority of these subjects ended up choosing the rolling strategy, a substantial proportion used static rehearsal, and still others reported a hybrid strategy.

In order to investigate the effects of strategies on performance, I separated the non-trained subjects by the strategies they reported. For convenience I limited the analysis to subjects who unambiguously chose either the rolling or static strategies, which accounted for 25 of the 31 control subjects. Of these, 10 used the static strategy and 15 used the rolling strategy.

An ANOVA performed on control (non-trained) subjects found, as can be seen in figure 5.6, that subjects who chose the static strategy were better able to discriminate between match and non-match trials, $F(1,21) = 8.95$, $p < .01$, $\eta_p^2 = .29$. The full ANOVA can be found in Appendix D (table D.9) Conclusions based on the full dataset comparing subjects trained to use the rolling strategy to those not trained needs to be updated in light of this finding of a sensitivity difference related to chosen strategy among non-trained subjects. The ANOVA of the full dataset found a marginal effect of strategy, but the analysis of control subjects suggests that subjects who chose the static strategy were as accurate as the trained rolling subjects. Interestingly, this means that subjects trained to use the rolling strategy were in fact more sensitive than non-trained subjects who chose the rolling strategy for themselves. The analysis revealed no other significant main effects or interactions for sensitivity (d') or criterion (c) for non-trained subjects.

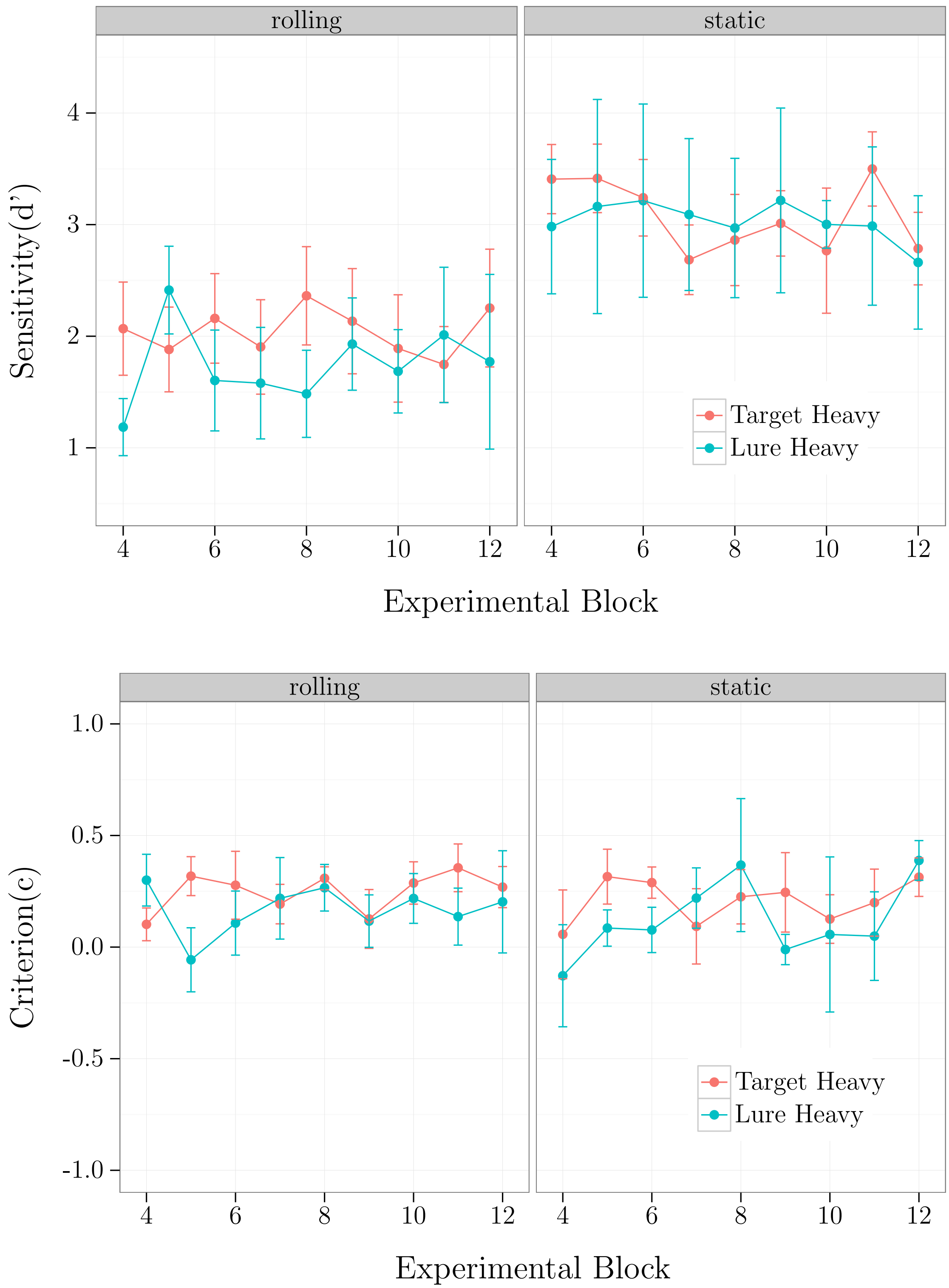


Figure 5.6: Plots of sensitivity (top) and criterion (bottom) by block for non-trained subjects. Sensitivity is shown in units of d'. Higher values of d' represent an improved ability to distinguish between match and non-match responses. Criterion values range from 1 to -1, with higher values relating to a bias to respond no. Error bars represent standard error.

In addition to responding more accurately, subjects who chose the static strategy (M = 793.21 ms, SD = 46.48 ms) also responded more quickly than subjects who chose the rolling strategy (M = 875.67 ms, SD = 42.02 ms), As figure 5.7 shows, static subjects were faster at all lag positions for both target heavy and lure heavy sequences.

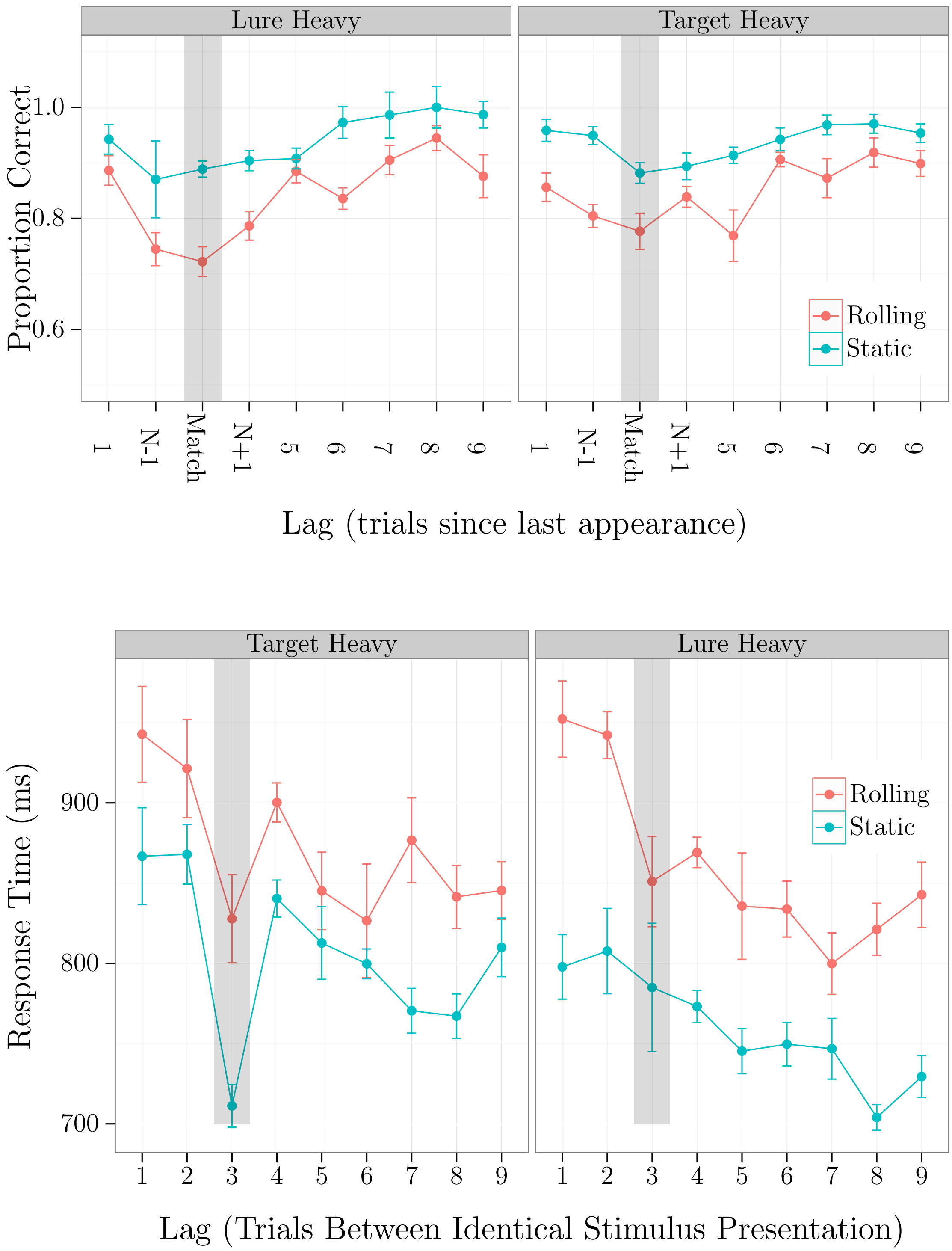


Figure 5.7: Plots of response time by lag position for non-trained subjects. The shaded area is lag position 3 (match trials). Error bars represent standard error.

As with experiment 1 and the full dataset of experiment 2, I investigated whether strategy differences may have led to differences in control mode by calculating mean response time differences between match and lure trials. An anaylsis of variance found a significant main effect of T:L ratio $F(1{,}21) < .01$, $\eta_p^2 = .27$. Although Figure 5.8 suggests that the static strategy leads to larger differences for target heavy sequences than does the rolling strategy, the ANOVA did not find a significant interaction ($p = .22$). The small number of static subjects is likely to blame for the lack of a significant effect. The effect size for this interaction ($\eta_p^2 = .06$) suggests that a significant interaction might be found with a larger sample size.

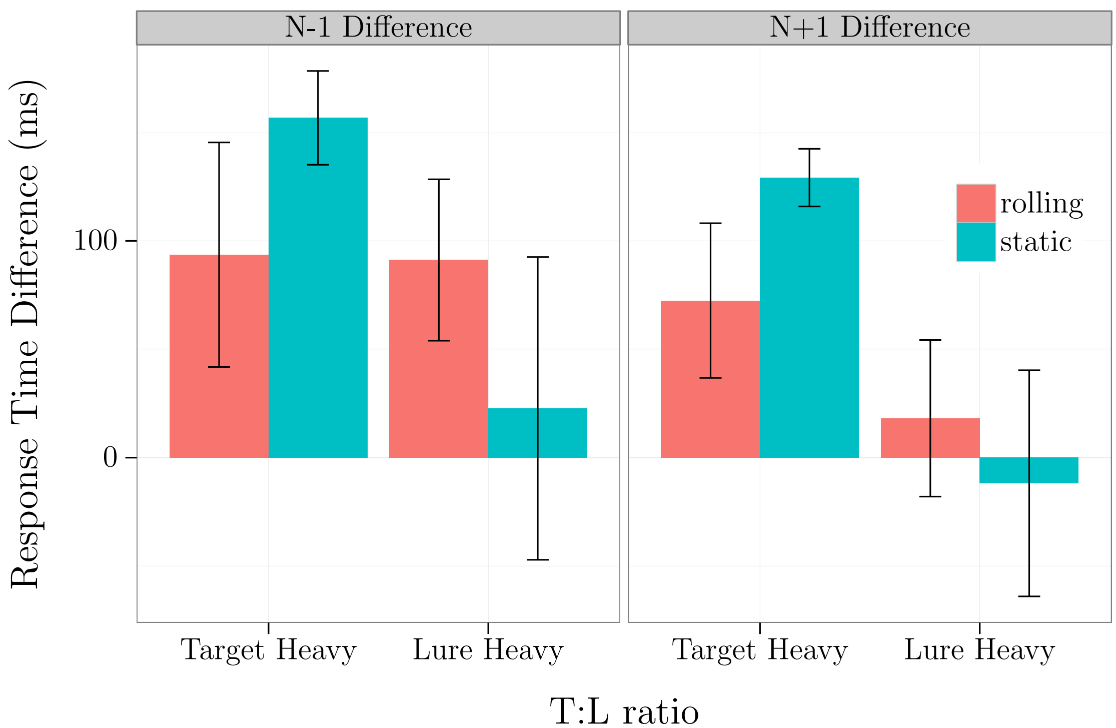


Figure 5.8: Differences between lure and match trial response times. Left: Difference between n-1 lure and match trials. Right: Difference between n+1 lure and match trials. Error bars represent standard error.

5.4 Discussion

5.4.1 T:L Ratio

As discussed previously, the results from experiment 1 failed to support the hypothesis that lure heavy sequences would bias people toward a proactive strategy. In fact, accuracy and response time data suggested that these sequences may have led to a reactive strategy. This was a surprising result given that research with multiple paradigms supports the hypothesis of increased proactive control as the likelihood of conflict increases. But unlike these other control paradigms such as the Stroop and flanker tasks, conflict in the n-back varies by response. "Yes" responses are generally congruent with respect to familiarity and activation, while "no" trials (especially lures) are incongruent. In the Stroop and flanker tasks, by contrast, conflict is unrelated to the actual response. For instance, a "green" trial in the Stroop task can be either congruent or incongruent, so subjects could not develop a strategy based on the choice of response like they seem to have done in the lure heavy conditions of experiment 1.

It is possible that the constant feedback provided in experiment 1 could have made the lack of "yes" trials more salient than would otherwise have been the case. Experiment 2 used the same sequences as experiment 1 but removed all feedback. Without constant feedback, subjects should have been less able to determine the relative lack of match trials. When compared with the signal detection results from experiment 1, the experiment 2 results (figure 5.2) suggest that the lack of feedback reduced the bias to respond "no". Criterion values increased over time for lure heavy sequences only when subjects were explicitly trained to use the rolling rehearsal strategy. Given that these subjects learned over time to match their performance to the nature of the task environment, it is possible that the rolling strategy may be more sensitive to reactive factors in the environment than is commonly assumed in the literature.

Contrary to experiment 1, lure heavy subjects in experiment 2 did use some form of proactive control, evidenced by the plot of response time by lag (figure 5.5). The amount of control used was greater for target heavy sequences, but this effect was not found at all for lure heavy subjects in experiment 1. These data suggest an influence of trial by trial feedback on the control strategy used. As discussed previously, this likely reflects the relative salience of the reward structure of the environment. When provided with constant feedback (as in experiment 1), subjects learned to use reactive control, responding yes only when they were absolutely sure an letter was a match. Without constant feedback, subjects were less able to determine the true ratio of yes vs no responses in a block of trials and did not make as strong an adjustment toward reactive control.

Combined, the results of the first 2 experiments suggest a complex relationship between T:L ratio and control mode. Instead of reacting to the likelihood of conflict (ratio of matches to lures), people seemed to react instead simply to the likelihood of match trials appearing, regardless of the existence of lure trials. In sequences where match trials were more common, subjects used proactive control in order to be able to recognize them. But in sequences with relatively few matches, the added cognitive effort may not have been needed to obtain satisfactory accuracy rates. However, the ability to make these control adjustments relies heavily on the ability to extract statistical information from the environment. When not made aware of the relative rarity of match trials, subjects used proactive control to keep the matching item active, as hypothesized.

This new hypothesis could be studied in future research by using sequences which hold the number of match trials constant but vary the likelihood of lure trials and the existence of trial by trial feedback. For instance we could compare sequences with 20 match trials and 10 lure trials to sequences with 20 match trials and 30 lure trials. These types of sequences could be created by simply reducing the number of possible stimuli. For instance, if we use only 5 possible stimuli (and hold the number of match trials constant), the proportion of lure trials will naturally increase.

5.4.2 Strategies

Taken as a whole, the results comparing the trained subjects to their control counterparts was underwhelming. No significant differences existed between trained and control subjects on either accuracy (as measured by d') or response times. In retrospect, the lack of significance should have been expected because the control groups do not represent the lack of a control strategy, but rather the lack of a *trained* control strategy. As self-reports made clear, these subjects did in fact use a control strategy and a majority of them used the rolling strategy. Thus significant differences should not have been expected.

Analysis of control subjects by selected strategies yielded more interesting results. Subjects choosing the static strategy were faster and more accurate than those using the rolling strategy. The subjects trained to use the rolling strategy outperformed non-trained subjects who chose the rolling strategy. The notion that the rolling strategy represents an ineffective proactive strategy fits with the response time data discussed above in which subjects trained to use the rolling strategy experienced more conflict on n-1 lure trials (see figure 5.4) Given these results it is clear that strategy differences do affect the ability to perform tasks like the n-back.

5.4.3 General Discussion

The results of the first 2 experiments yield evidence of a rich connection between the statistics of the environment and control mechanisms used to perform the n-back. People performing these tasks are sensitive to the stimulus distribution, T:L ratio, and proportion of matches to non-matches. Some of these effects seem to involve the salience of the reward structure (e.g. the effect of feedback on control mechanisms for lure heavy sequences), while others likely operate on the cognitive mechanisms that underlie performance, such as memory activation values in the case of stimulus distribution manipulations

Individual differences in strategy choice also affected performance, as subjects discovering the static strategy outperformed their rolling counterparts. The rolling strategy, the dominant strategy selected by subjects and discussed by researchers, seems to be an extremely ineffective proactive strategy. Although the rolling strategy surely involves proactive activity, that activity does a poor job of isolating and maintaining the target item. When a stimulus is presented, it is likely that the rolling strategy still requires a controlled search of memory to determine the correct response. The static strategy, by contrast, specifically activates and maintains the target item through the use of a pointer separate from the memory structure itself.

Experiment 3 will further investigate the difference between the rolling and static strategies by training groups of subjects in each strategy.

Chapter 6
Experiment 3

6.1 Purpose

Experiment 1 suggested a difference in rehearsal strategy caused by variations in stimulus frequency distributions. Lag position interacted with stimulus distribution, leading to an increase in response times on n-1 lures compared to match trials. Only even sequences exhibited this effect and response time differences persisted across T:L ratios.

An increase in response times for n-1 lures is evidence of a difference in rehearsal strategy (or at least in the amount of rehearsal). I expected slower response times on n-1 lures because these letters are in the current rehearsal window, and excessive rehearsal increases the activation of n-1 lures, leading to slower response times. Accordingly, Szmalec found an increase in the difference between match and n-1 lure trials due to ISI length such that longer ISIs led to slower n-1 lure response times (Szmalec et al., 2011). Szmalec concluded that the increase in ISI allowed time for more rehearsal, which increased the activation of the n-1 lure stimuli. The results of experiment 1 demonstrate that factors other than ISI can lead to similar effects. But the finding of a response time difference does not in and of itself answer the important question of why skewed and even sequences differ in the type or quality of rehearsal they stimulate.

Experiment 3 was designed to test two competing hypotheses that explain this effect. One hypothesis is that people rehearse less when the same items appear frequently over a short time span, as occurs in our skewed sequences. These items remain highly active and are easier to remember when presented as matches. It is possible that subjects recognize this feature of skewed sequences and choose to rehearse less because these items are more salient. This feature of skewed sequences was discussed in detail in chapter 2. People might need to rehearse more when presented with even sequences because they are not being "helped" by the statistics of the environment. Even sequences are more evenly distributed and thus do not include excessive repetitions of specific items. As a result, successful discrimination of matches may be more dependent upon rehearsal for even than for skewed sequences. Response time and accuracy data from experiment 1 support this explanation. Accuracy was lower and response times longer for n-1 lures in the even sequences when compared to skewed sequences. Further support was provided by the finding that response times for n+1 lures, which were not in the current rehearsal window, did not differ between even and skewed sequences.

A second hypothesis is that stimulus distribution affects the selection of control strategy. People may be more likely to use the static strategy when presented with items that tend to be repeated over short time periods, such as is the case for skewed sequences. Because the static strategy does not involve re-ordering the memorized list of the past three letters on each trial, it may be preferred when stimulus repetitions are more common. With regard to the first hypothesis, the static strategy likely requires less rehearsal to keep the current list of items active. Only one of the three items is changed on each trial. As the instruction slides for the static strategy in the appendix demonstrate, each item stays in the same sequential location in memory for three consecutive trials. By contrast, the rolling strategy requires a new position-item binding for each item on each trial. Because of this critical difference between strategies, it serves to reason that the rolling strategy would require more rehearsal to keep the last three items active in memory (in the correct order).

As in experiment 2, researchers trained subjects in the use of a specific strategy. In this case subjects were trained to perform either the rolling or static strategies. A third no-training group was included to determine if a difference in strategic preference develops as a result of changes in stimulus distribution.

6.2 Methods

6.2.1 Participants

144 subjects were recruited from the RPI campus and were awarded course credit in exchange for their participation. Two subject were removed from the analysis because of equipment malfunctions.

6.2.2 Materials

Experimental sessions followed the procedure of experiment 2 with a few exceptions. Unlike experiment 2, a small black box appeared in the center of the screen if no response was made within 2000 ms of stimulus onset, signaling to the subject the need to respond more quickly. This box remained on-screen for 1000ms. The black box was incorporated to encourage subjects to respond as quickly as possible.

6.2.3 Design

This experiment followed a 3 (strategy) x 2 (stimulus distribution) x 12 (block) design with strategy and stimulus distribution manipulated between subjects. Each subject completed a 3 mini-block training procedure followed by 12 experimental blocks. [3] Subjects were randomly assigned to one of six between-subjects conditions. The conditions span the three strategies (Rolling, Static, no-training control) and two types of sequences (even or skewed). Researchers originally planned to use 20 subjects per condition, but added extra no training control subjects to increase the power of a comparison of these subjects chosen strategies. These subjects were added at the end of the experiment. Figure 6.1 shows the resulting counts of subjects by condition after removing all subjects who did not complete the full experiment.

Table 6.1: Conditions for Experiment 3

Condition	Targets	Lures	Stimulus Distribution	Subjects/Condition
Rolling-Even	24	8	Even	21
Rolling-Skewed	24	8	Skewed	20
Static-Even	24	8	Even	19
Static-Skewed	24	8	Skewed	20
No-training-Even	24	8	Even	31
No-training-Skewed	24	8	Skewed	31

All sequences included 24 match trials, 4 n-1 lure trials, 4 n+1 lure trials and 32 non-lure distractor trials. Unlike experiments 1 an 2, subjects were presented with

[3] As in Experiment 2, subjects were given the option of completing a forth practice block if they did not yet feel comfortable with the task or assigned strategy. 33 of the 142 subjects chose to complete the optional practice block

sequences randomly selected from a pool of 1000 pre-generated sequences for each condition (even or skewed). The experimental software randomly chose 12 numbers between 1 and 1000 for each subject. The 12 sequences used in the experiment corresponded to these random numbers. Thus each subject was presented with a different set of 12 64 trial sequences. Each subject was trained in one strategy using randomly selected sequences of the same type the subject would experience during experimentation. The training sequences use the first 32 trials of three sequences randomly selected from the same list of 1000 sequences used in the experimental portion of the session.

6.2.4 Procedure

Although experiment 3 employed a similar training procedure as was used in experiment 2 (with the addition of the static strategy), it is included here in detail to avoid any confusion. Each experimental section began with a training procedure. All subjects were instructed on the basic principles of the n-back task. Researches urged subjects to focus on performing the instructed strategy regardless of resulting performance and even if they feel there was a better way to do it. Control subjects were given no strategy related information.

Researchers then gave detailed instruction to subjects on the use of one strategy (rolling, static, or no-training). Presentation slides for this instruction are included in the appendix. Subjects in the no-training group received only rudimentary instruction consisting of a basic description of the task (e.g. determine if the current letter is the same or different than the third previous letter). The researcher then talked the subject through a sample 7 trial sequence to make sure they understood the basic rules of the task, but presented no strategy related information.

All subjects (including the control group) performed three 32-trial practice blocks. The experimenter remained in the experimental room during these training blocks. These sequences were randomly selected for each subject from the list of 1000 pre-compiled sequences matching the stimulus distribution of the current experimental condition. The first 32 trials of these sequences were used for practice. The first practice block was self-paced with feedback. Subjects could take as much time as necessary to make responses. The lack of time pressure gave the experimenter a chance to reinforce the trained strategy and confirm that the subject was in fact using the strategy correctly.

Subjects in the strategy groups (rolling and static) were asked to report the current list of the previous three items in the correct order out-loud after each letter appeared so that the researcher could confirm that they were using the correct strategy. For the rolling strategy this rehearsal list was the last three items presented in sequential order (e.g. BCD if the last 4 were A,B,C,D). For the static strategy, the correct order depended on the current position of the sequence pointer. If the last 4 letters were A,B,C,D, and the last position of the pointer was 1, then the subject should report DBC. If the pointer was "2", then the subject should report ADC. As in experiment 1, a colored box appeared after each response. A red box indicated an error and a green box followed a correct response.

The second practice block also provided feedback via colored boxes but was not self-paced. Instead, the second practice block followed the timing of the experiment (letters appeared with a 3000 ms SOA). This block provided subjects with the opportunity to get accustomed to the time pressure of the study while still practicing the instructed strategy. The third practice block also followed the standard timing of the experiment, but did not provide feedback after each response. This practice block gave subjects a final chance to practice the strategy and prepare for the experimental blocks, which except for the difference in length are identical in design (no feedback, 3000 ms SOA).

At the end of the 3rd practice block subjects in the strategy conditions were asked to rate their ability to perform the instructed strategy on a scale of 1 to 7. They were then be asked if they would like to perform an additional practice block. If chosen, the additional practice block would be identical in design to the 3rd practice block. 33 of the 144 subjects chose to complete this 4th practice block.

The experimental phase of the session was divided into 12 blocks, with each block including one 64-trial sequence sampled from 1000 pre-compiled sequences. As in experiments 1 and 2, all sequences sampled from the following 8 letters: B,F,H,J,M,R,X,Z. All sequences included 24 match trials, 4 n-1 lure trials, 4 n+1 lure trials, and 32 non-lure distractor trials. Each sequence in the even conditions was derived using the discrete optimization algorithm discussed in chapter 3 and include 8 of each of 8 possible letters. To reiterate, there was no variability in the distribution of stimulus frequencies between these sequences.

Skewed sequences were derived by randomly selecting stimuli to match a pre-determined trial type ratio (3:1). This method, common in the literature, produces sequences with skewed stimulus distributions. As described in chapter 2, the most common 2 or 3 items often comprise over 50% of the trials in these sequences. The python code for this algorithm is included in the appendix. This method produces a range of stimulus distributions. In order to investigate the effect of stimulus distribution, it was important that the sequences chosen for the skewed conditions have as little variability as possible. To this end, 20,000 sequences were pre-compiled, and only the most skewed sequences were selected for inclusion in the final 1000 pre-compiled sequences. The measure of stimulus distribution used to select sequences was the sum of the number of trials occupied by the three most common letters chosen for each sequence. Of the original 20,000 sequences, the three most common letters occupied a mean of 35.41 out of 64 trials (compared to 24 trials for the even sequences) with SD=2.91 trials. The three most common letters for the 1000 skewed sequences chosen for this experiment occupied a minimum of 40 trials. The top 3 items in these 1000 sequences occur on a mean of 41.09 trials with SD = 1.3 trials.

Following the experimental session subjects were asked several questions based on the experimental condition. Subjects in the strategy conditions were asked to rate their success at performing the instructed strategy. They were also asked if they found themselves using a different strategy and to describe the alternate strategy. After describing an alternate strategy, the researcher determined if the described strategy matched either the static or rolling rehearsal strategies or a hybrid version. In order to do this, the experimenter presented the subject with a hypothetical subsequence and asked the subject what they would do next (e.g. "if the last three letters were R, D, F, and the new letter is a C, what would you do next"). If the subject indicated that her new memory list would be D,F,C, the experimenter coded the strategy as rolling. If the subject replied C,D,F, the experimenter coded the strategy as static. Using a similar procedure, researchers asked no-training subjects to describe their memory strategy used in their own words. Again using the procedure described above, the experimenter coded the strategy as rolling, static, or hybrid.

A small number of subjects in experiment 1 reported using a strategy that could not easily be categorized as static or rolling. This generally consisted of the memorization of letters in blocks of three. Each new item was compared to the memorized list in much the same way as is done in the static strategy, but after three items the old list was discarded and the new list of the past three was recompiled from memory. For instance if the past three items were A,B,C and the next three were D,B,A, then the subject would compare each letter sequentially (D!=A, B=B, A!=C) and then rehearse DBA as the new comparison list for the next three trials. This strategy was coded as a hybrid strategy as it combined elements of the static and rolling strategies. Researchers then asked if the subject settled on a strategy early in the experiment or if they experimented with other strategies as well. I performed subsequent analyses of non-trained subjects with respect to their chosen strategies. These analyses were limited to subjects who indicated using either the rolling or static strategies.

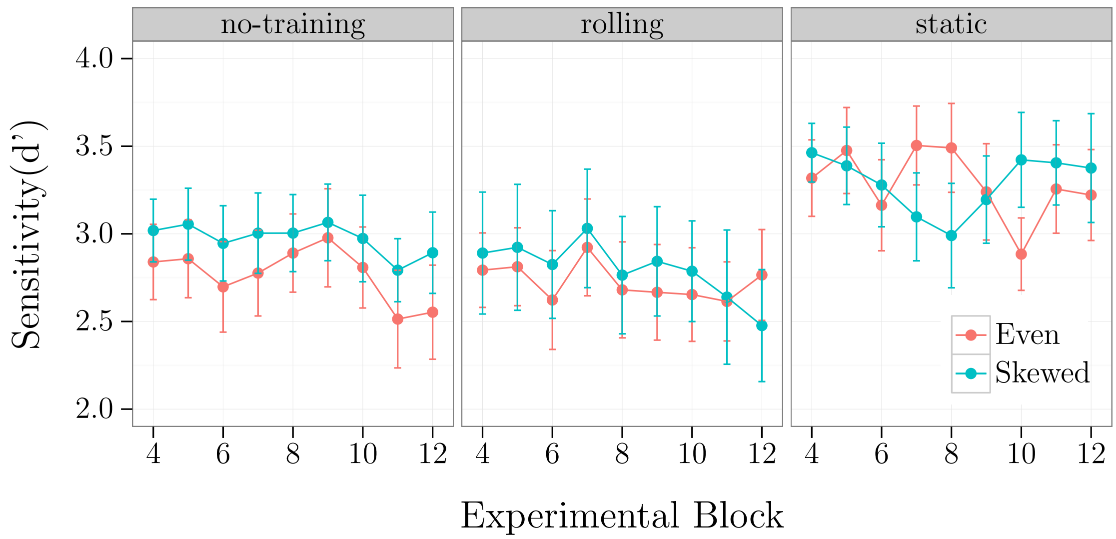
6.3 Results

Like experiment 2, analyses were performed separately on the omnibus dataset and the subset of no-training subjects. Analyses were again designed to be comparable between experiments, so all methods for calculating signal detection, and response time differences were identical to those used in previous experiments.

6.3.1 Response Accuracy

Like experiments 1 and 2, I tested for linear trends in both the full and truncated datasets for sensitivity and criterion. A significant quadratic trend by block in the full dataset was eliminated by truncating the initial 3 experimental blocks. After truncating the dataset a significant linear trend of block remained, but visual inspection of figure 6.1 suggests that any linear effect reflected a decrease in sensitivity toward the end of the experimental session, possible indicating fatigue. Regardless, this does trend does not relate to the hypotheses of this work. Having removed the significant quadratic term by truncating the data, and to provide consistency throughout the three experiments, all further analyses were performed on the truncated dataset.

ANOVA and post-hoc comparisons of sensitivity (d') revealed that subjects in the static conditions (M = 3.32, SD = .71) were more sensitive than those in the rolling (M = 2.79, SD = .99) and no-training (M = 2.84, SD = .94) conditions, $F(2,136)=3.21$ $p<.05, \eta_p^2=.04$. Post hoc comparisons revealed that the differences between the static and rolling ($t=2.62$, $p<.01$) and no-training ($t=2.77$, $p<.01$) conditions were the only significant comparisons. There were no other significant factors influencing sensitivity. These results are displayed in Figure 6.1 (top panel). There were no significant main effects of strategy or stimulus distribution on the criterion value.



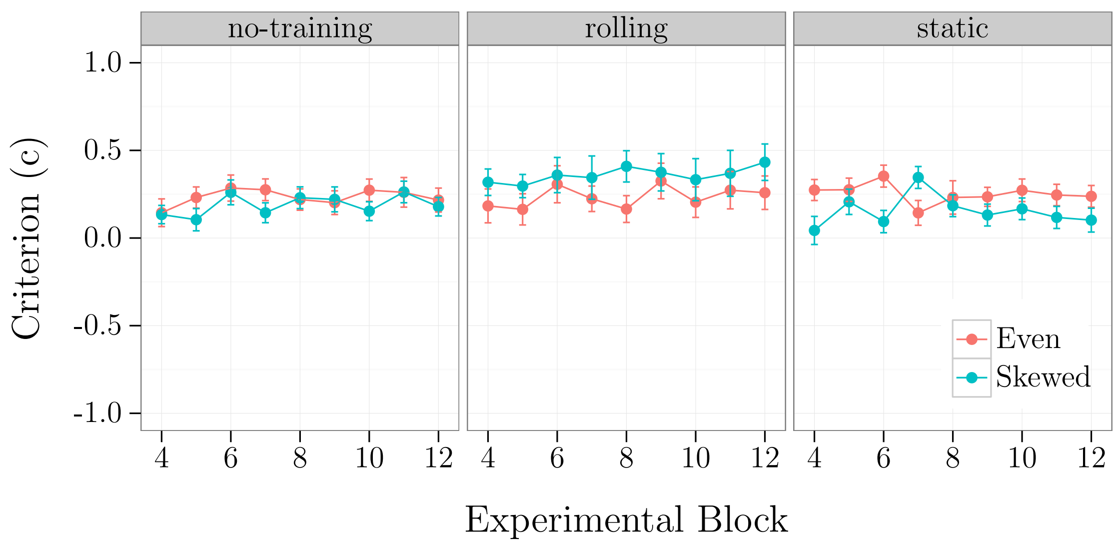


Figure 6.1: Top: Sensitivity (d') over time. d' measures the difference in units of standard deviation between correct responses on match trials and false alarms on non-match trials. Bottom: Criterion (c) is a measure of response bias. Values closer to 1 denote a bias to respond "no", Values closer to -1 denote a bias to respond "yes". 0 represents no bias. Error bars represent standard error.

Figure 6.2 displays accuracy rates by position lag and is a bit more informative as to the differences caused by the strategy manipulation. The static strategy outperformed the rolling and no training strategies at all lag positions, but especially at the critical match and lure positions. In fact the accuracy rates for static subjects were the highest recorded when compared to all conditions of all three experiments.

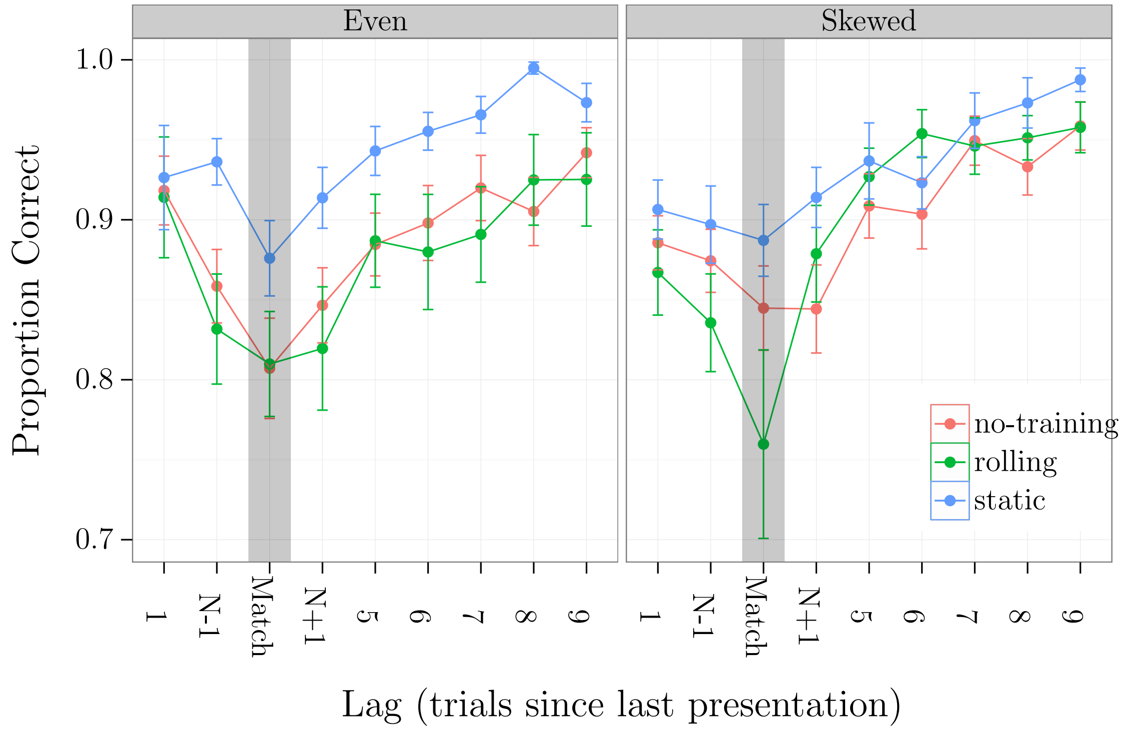


Figure 6.2: Accuracy by lag position. Shaded box denotes match trials. Error bars represent standard error.

6.3.2 Response Time

Response time differences between match and lure trials were calculated in the same manner as in experiments 1 and 2. Figure 6.4 is a plot of response time differences (with respect to match trials) for n-1 and n+1 lures, respectively. A 2 (distribution) x 3 (strategy) x 2 (lure type) ANOVA found main effects of stimulus distribution $F(1,136)=23.31$ $p<.001,\eta_p^2=.13$, and strategy $F(2,136)=6.96$ $p<.002,\eta_p^2=.09$. Stimulus distribution also interacted with lure type $F(2,136)=23.13$ $p<.001,\eta_p^2=.04$. Stimulus distributions affected the RT difference more strongly for n-1 than for n+1 lures. This result replicates the complementary analysis of response times in experiment 1, strengthening the claim that stimulus distributions relate to differences in control strategies which result in differential effects on n-1 vs n+1 lure trials.

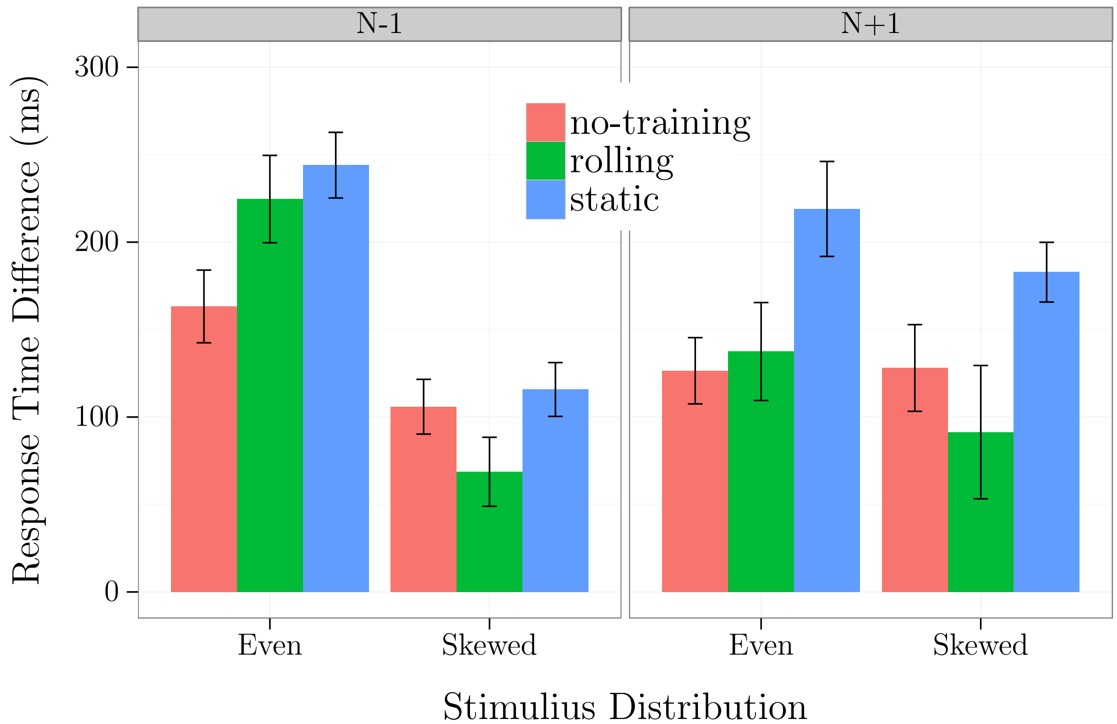


Figure 6.3: Response time differences between match and lure trials. Left panel represents RT differences between n-1 lures and match trials. Right panel represents differences between n+1 lures and matches. Error bars represent standard error.

The increase in n-1 RT differences for even vs skewed sequences replicated the results for experiment 1. Likewise, an analysis of n-1 response times by lumpiness replicated the experiment 1 finding of reduction in response time with increasing lumpiness for skewed sequences, but no such effect for even sequences. An analysis of covariance found a stimulus distribution by lumpiness interaction, $F(1{,}658)=9.03$ $p<.003, \eta_p^2=.01$. The co-variation between response time and lumpiness for skewed sequences occurred for all three strategy conditions. This data does not show any difference attributable to strategy selection (or training), thus stimulus distribution seemed to directly affect basic control mechanisms, regardless of chosen (or trained) strategy.

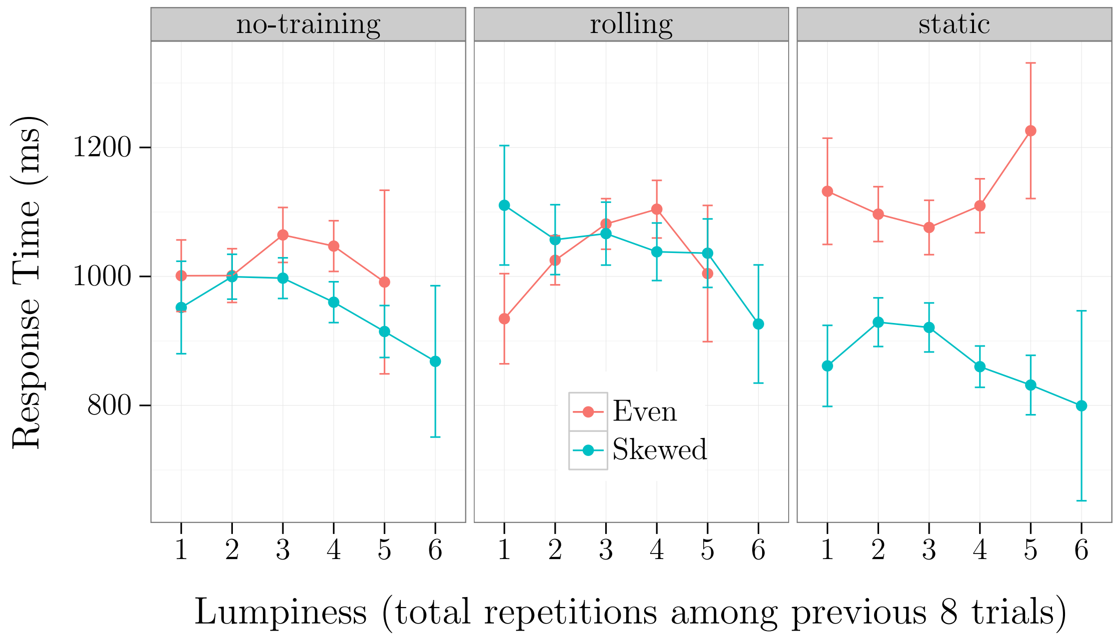


Figure 6.4: Response times by lumpiness for n-1 lure trials. Error bars represent standard error.

A plot of response times by position lag (Figure 6.5) revealed several hard to explain effects. First, static subjects were faster than rolling subjects when presented with skewed sequences. But static subjects were slower than rolling subjects for even sequences. Also, subjects not trained matched the response time curves of rolling subjects for even sequences and static subjects for skewed sequences.

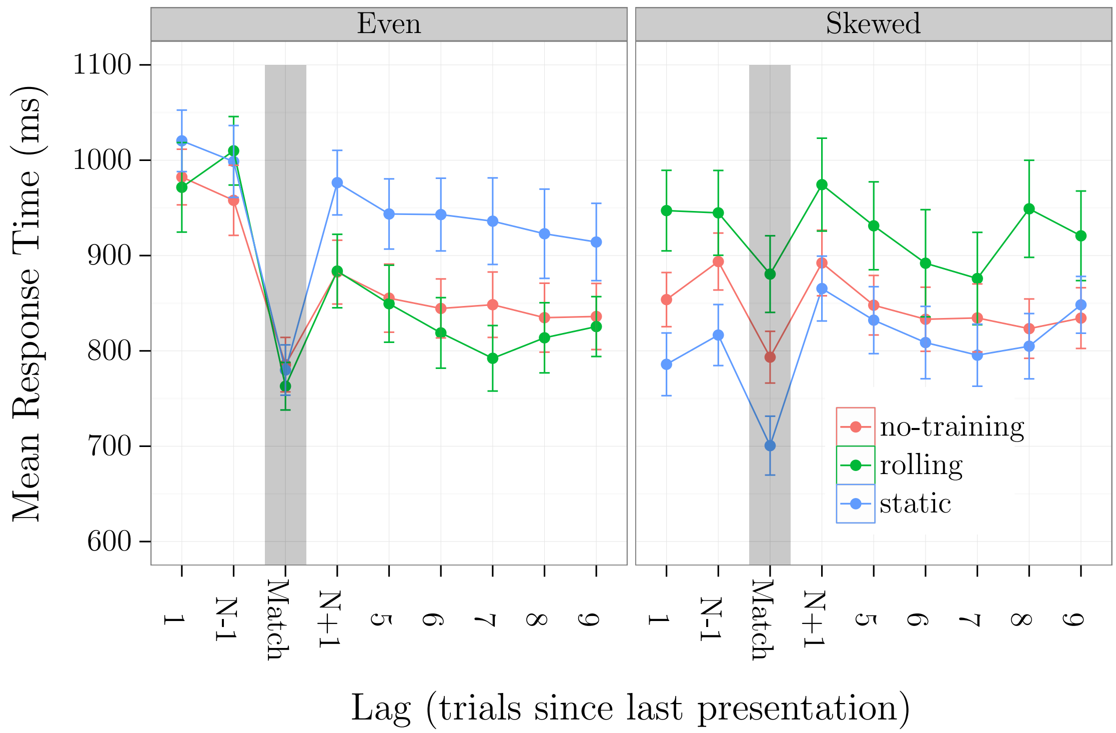


Figure 6.5: Response time by lag position. Error bars represent standard error.

6.3.3 Non-trained subjects

Results from Experiment 2 suggested effects of self chosen strategies on various metrics of performance, but in there was insufficient power to make strong conclusions. The power of these analyses was increased in experiment 3 by adding extra subjects to the non-trained conditions. Experiment 3 included a total of 62 non-trained subjects, over 50% higher than were used in experiment 2.

The response time by lag plot for the current experiment (figure 6.5) suggested that non-trained subjects may have been more likely to choose the static strategy for skewed sequences and the rolling strategy for even sequences. Interestingly, there was in fact no difference in the number of subjects choosing the static strategy between even and skewed sequences. Of those who selected either static or rolling, 10 of 28 in the skewed condition and 11 of 27 in the even condition chose the static strategy.

As in experiment 2, these analyses were limited to subjects who unambiguously chose either the rolling or static strategies. There were no significant or marginal main effects or interactions for either sensitivity or criterion. Although there were no significant differences in sensitivity, figure 6.6 does suggest differences in accuracy by lag position related to strategy choices. In fact, an ANOVA of accuracy rates by lag revealed an interaction between chosen strategy and lag position $F(9{,}450) = 2.67$, $p<.005$, $\eta_p^2 = 0.06$. The full ANOVA can be found in table E.8 in the appendix. As can be seen in figure 6.6, subjects choosing the static strategy were more accurate on target and lure trials. These data replicate the finding of effective proactive control for the static strategy when compared to rolling. Response times did not differ by strategy, and there was no interaction of strategy and stimulus distribution as was found for the trained subjects.

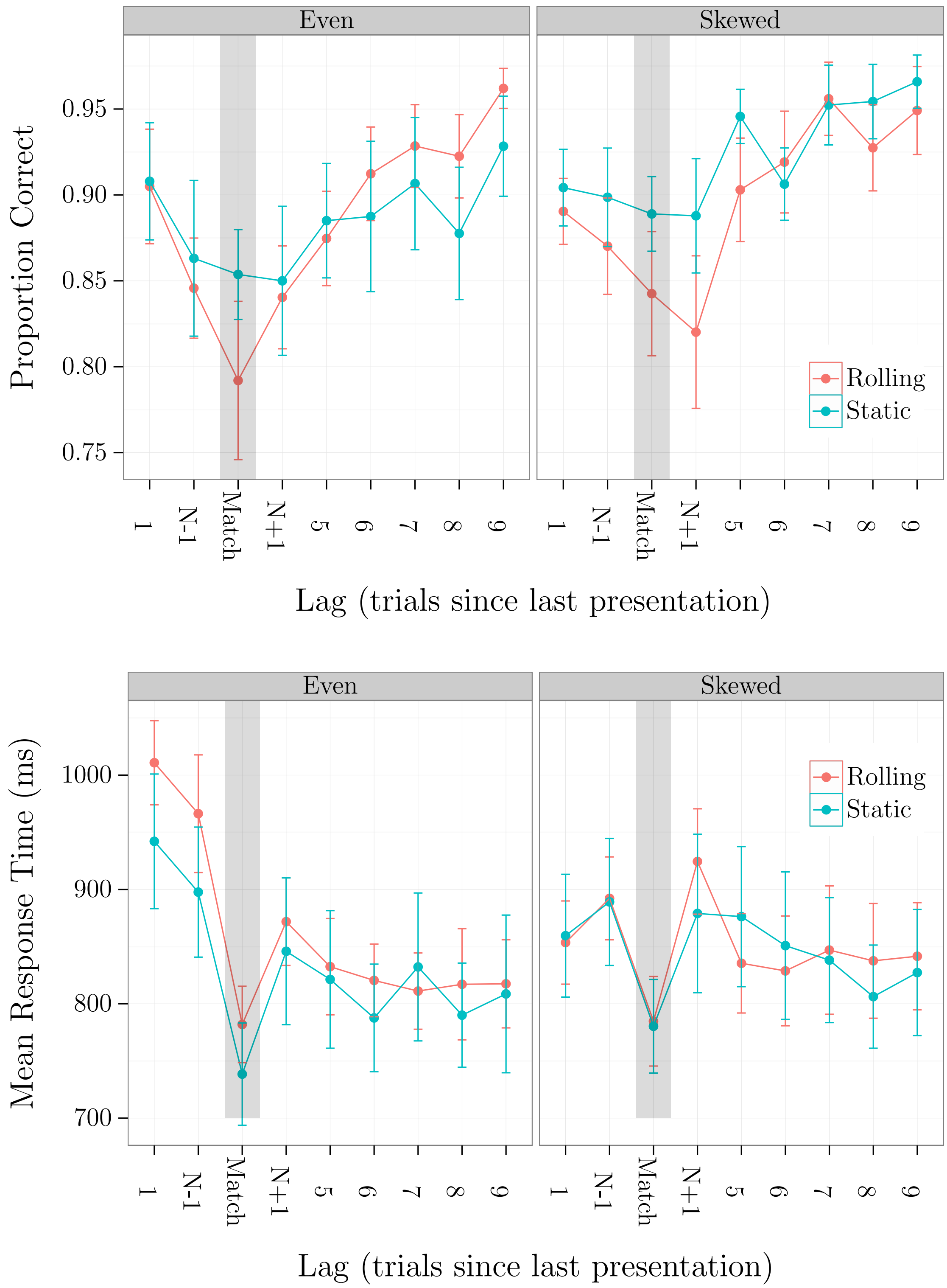


Figure 6.6: Mean accuracy and response times by lag position for control subjects. Error bars represent standard error.

An analysis of lure-match RT differences found a significant stimulus distribution by lure position interaction $F(1{,}50) = 8.67$, $p<.005$, $\eta_p^2 = 0.14$., but no effects related to strategy choice. The distribution by lure position interaction is another replication of the same effect found when examining the full dataset of experiment 3 and the data from experiment 1. In all three instances even sequences resulted in larger differences between lure and match RTs, but only for n-1 lures. As discussed previously, this effect suggests that even sequences lead to increased rehearsal of the past three items, while skewed sequences lead people to become more sensitive to local lumpiness, adjusting rehearsal rates accordingly.

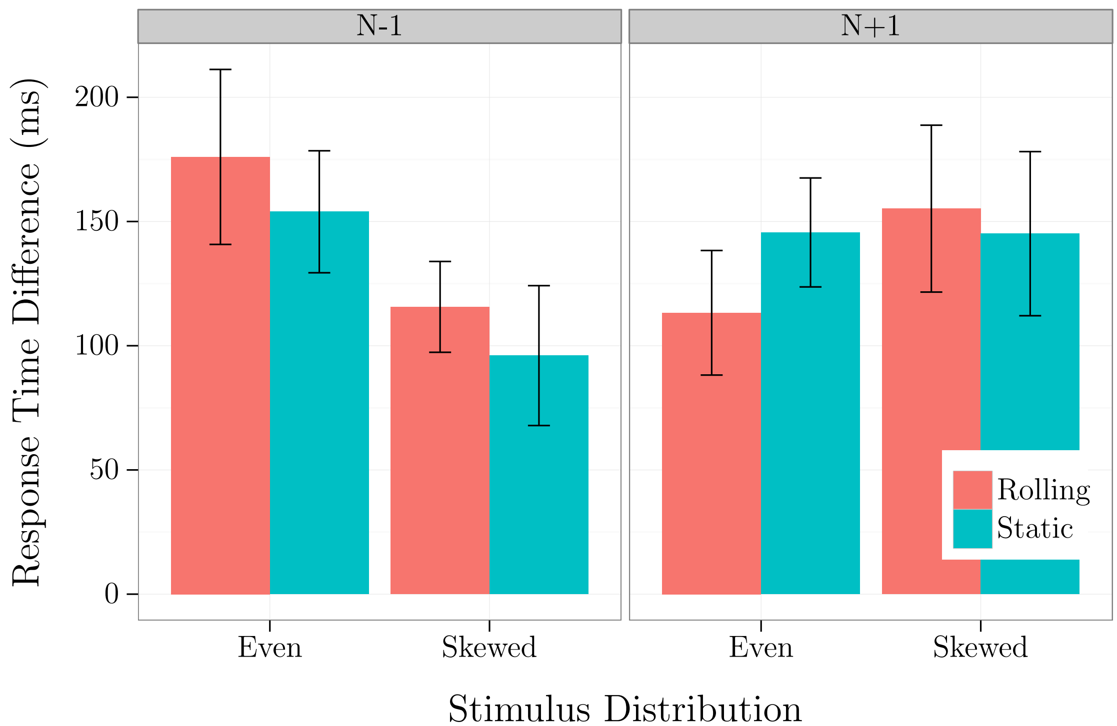


Figure 6.7: Response time differences by lure type. Error bars represent standard error.

6.4 Discussion

Experiment 3 aimed to determine how stimulus distributions and strategy selection interact to effect the use of proactive control in the n-back. Experiment 1 found a significant effect of stimulus distribution on response time measures, and the results of experiment 3 replicated that effect. Subjects in the even conditions had larger differences in response times between target trials and n-1 lure trials regardless of whether they were trained in the rolling, static, or were not trained.

Performance of subjects trained to perform the static strategy significantly differed from the other two groups. Static trained subjects responded more accurately than the other groups across all lag positions, and the performance differences reached their peak at the critical lure and target positions. These accuracy differences persisted across both stimulus distribution conditions. This difference was attenuated for control subjects. Control subjects choosing the static strategy were significantly more accurate on match and lure trials but not for distractors. It is possible that these subjects did not choose the strategy until late in the experiment, or indicated they used rolling rehearsal when they actually used the static strategy or a hybrid strategy.

Response time results suggest that the accuracy improvement for static trained subjects may be related to the relative use of proactive control. Static trained subjects exhibited larger RT differences for n+1 lure trials, providing evidence that the static strategy involves stronger maintenance of the target item. As discussed previously, n+1 lure trials feature items that were the target item on the previous trial, so an increase in the RT difference between these trials and match trials is consistent with the hypothesis that extra activation of the target item leads to increased conflict on the next trial. In the case of static trained subjects, this extra activation may have reduced errors for both lure and target trials, as can be seen in figure 6.2.

As for the rolling strategy, the results of experiment 3 add to a growing list of evidence which suggests that it is a particularly inefficient control strategy. The rolling trained subjects performed no better than subjects who were not trained. Additionally, RT difference measures suggested an increase in rehearsal (n-1 lure RT difference) but no extra activation reserved for the target item (smaller increase in n+1 lure RT difference). Lack of focus is one possible explanation for the relatively poor performance of rolling subjects, but the emphasis placed on performing the rolling strategy should have increased the focus of the trained subjects. The fact that there was no discernible difference in performance between trained and non-trained rolling subjects makes it questionable to blame lack of focus. Thus these results, combined with those of experiment 2, point to the control features of the rolling strategy as the source of errors.

The analysis of non-trained subjects provided mixed results. Accuracy data replicated the effect found for the omnibus dataset, with subjects choosing the static strategy outperforming rolling subjects at the critical n-1, match, and n+1 lag positions. Response time differences showed no difference related to strategy (but replicated the effect of stimulus distribution found several times in the current work).

I'll close this discussion by focusing on a fascinating set of results that may provide fertile ground for future research. Figure 6.5 shows a strong interaction between trained strategy and stimulus distribution. Static trained subjects responded faster for skewed sequences than even sequences. Conversely, rolling trained subjects responded more quickly on even than skewed sequences. It is possible that this interaction provides evidence of differential control requirements of the two strategies, but more research is necessary to replicate and examine this effect.

Adding to the curiosity, non-trained subjects' response times matched the response times of rolling subjects on even sequences, but matched static RTs for skewed sequences. Non-trained subjects could be more likely to use the static strategy on skewed sequences, a hypothesis supported by the distribution of chosen strategies in experiment 1. However, there was no difference in the frequency of static self reports related to stimulus distribution. Of course it is possible that some subjects performed the static strategy without conscious awareness, but again, future research should investigate this effect.

A second explanation involves the selection of subjects to conditions. Given that a portion of subjects trained in one strategy would have naturally chosen a different strategy, some of the variance in the data could be related to training some subjects in their natural strategy and other subjects in a foreign strategy. By contrast, non-trained subjects were able to choose their natural strategy. As a result their response times showed no difference between even and skewed sequences.

Regardless of these questions, the interaction found for trained subjects does suggest that stimulus distribution affects the use of various control strategies.

Chapter 7
Discussion

7.1 General Discussion

The results of the three experiments described in this dissertation shed light on how manipulation of stimulus sequences can effect performance in the n-back task. More importantly, detailed analyses of response times revealed that sequence manipulation likely causes differences by affecting both the choice of cognitive control strategies and by affecting the control mechanisms that make up those strategies. As such, these studies pose important questions to researchers using the n-back to study working memory capacity. This concluding chapter will summarize the most important findings, discuss unanswered questions posed by the current studies, and suggest avenues of future research into this and related topics.

7.1.1 Summary of Results

The hypotheses of this dissertation concerned two related factors commonly manipulated in the n-back literature. Based on the cognitive control literature, I hypothesized that an increase in the likelihood of conflicting trials (lures) would lead to a bias toward proactive control strategies. I also hypothesized that the skewed sequences used in most n-back studies bias people toward a reactive control mode.

7.1.1.1 T:L ratio

The data from experiments 1 and 2 failed to support the hypothesis of increased proactive control for lure heavy sequences. In fact, the data strongly suggest that the lure heavy sequences used in the current study had the opposite effect, leading subjects to perform in a reactive manner. Subjects in the lure heavy conditions of experiment 1 developed a strong bias to respond "no", presumably because they learned that there were very few match trials per block (8 out of 64 trials), and developed a strategy that would maximize their ability to correctly reject non-match trials.

The removal of feedback (experiment 2) eliminated the bias found in experiment 1, and the response time data for lure heavy sequences showed evidence of proactive control. These data still indicated more proactive control for target heavy sequences. Still, the appearance of evidence for proactive control underscores the importance of experimental design in determining control mode. A detail as small as trial by trial feedback had a strong impact on the type of control subjects in these conditions used. The difference in response times between lure and match trials, my measure of choice for determining control mode differences, returned significant results with large effect sizes for both experiments 1 and 2 related to T:L ratio. Although the data from experiment 2 revealed some amount of proactive control, the effect size for the interaction between T:L ratio and lure position was practically equivalent in both experiments.(η_p^2=*.24 and .22, respectively*).

Subjects seemed to match control strategies to the reward structure of the environment instead of to the likelihood of conflict (lure trials in this case). When made explicitly aware of the likelihood of match trials via constant feedback, subjects made an optimal control choice. In environments with few match trials people prefer to act reactively because the effort required to keep the target item active is not worth the reward of correctly recognizing such a small number of matches. On the other hand, when presented with target heavy sequences, the effort required is worthwhile, as matches are much more common. Future studies should be conducted to test this hypothesis as will be discussed in detail later in this chapter.

7.1.1.2 Stimulus Distribution

T:L ratio manipulations in the literature are likely intentional. Some researchers may have wanted to avoid the biased responses I found in experiment 1 (Kane et al., 2007). Others may have required more match trials for statistical purposes. Stimulus distribution manipulations, on the other hand, almost certainly are not intentional, instead occurring as a side effect of random selection and T:L ratio manipulations. The results from experiments 1 and 3 strongly supported the hypotheses that skewed stimulus distributions commonly found in the literature do affect cognitive control in the n-back task. Although accuracy rates show little to no difference between sequence type, comparisons between skewed and even stimulus distributions revealed significant differences with large effect sizes for measures sensitive to the use of proactive maintenance of the target item. When presented with skewed sequences, the fact that some items are more likely to appear than others lessens the need to spend resources raising the activation of the target item. Even sequences provide no such help, so rehearsal and other proactive processes are necessary to make accurate comparisons with the nth back item.

To my knowledge the current work was the first to investigate the effect of within-sequence changes in lumpiness. Results concerning local lumpiness were consistent across experiments. Subjects in the skewed conditions of experiment 1 and 3 reacted to lumpiness, making faster responses to n-1 lures as lumpiness increased. Conversely, subjects in the even conditions of all three experiments showed no effect of lumpiness on n-1 response times. Although even sequences were less lumpy than skewed sequences, they still featured some variability in local lumpiness, often including subsequences with as many as 5 repetitions among 8 trials. The lack of a response to this factor of the environment suggests a strong proactive mode for subjects in the even conditions. By contrast subjects in the skewed conditions were responsive to changing lumpiness, likely changing the frequency of rehearsals to match the current activation values of recently seen items.

7.1.1.3 Strategies

Most researchers assume that proactive control in the n-back relates to sequential rehearsal of the last n items, identical to the rolling strategy discussed throughout this dissertation. However, the mechanisms of rolling rehearsal do not in and of themselves lead to effective proactive maintenance of the target item. Simply rehearsing the last n items in sequence raises the activation of all three items, making a memory retrieval necessary to determine if the current item matches the nth back item. Because of this reactive aspect to the rolling strategy, these sequences were associated with response time curves that were consistent with a reactive strategy. By contrast, the maintenance of position information in the static strategy is a form of proactive control, forcing the selective maintenance of the nth back item.

To my knowledge, experiments 2 and 3 were the first studies to explicitly train subjects in the use of particular n-back strategies. Both experiments showed that the rolling rehearsal strategy was an inefficient proactive strategy. Subjects trained to use rolling rehearsal were less accurate than those trained in the static strategy. In addition, response time measures suggested that rolling rehearsal led to smaller differences between match and n+1 lure trials, suggesting less use of proactive control. As discussed in chapter 1 the rolling strategy involves elements of reactive control because despite rehearsing the last three items, the preparation for the current item requires additional proactive control. For the most part, when a new stimulus appears, a memory retrieval is likely required to determine if a familiar item matches the third previous item. Data from subjects trained in the rolling strategy support the categorization of rolling rehearsal as at least a partially reactive strategy.

In contrast, subjects trained to use the static strategy in experiment 3 exhibited signs of effective proactive control. In addition to increased accuracy across all trial types, response times by trial type were consistent with predictions based on selective proactive maintenance of the target item. Specifically, increased response times on n+1 lure trials may have been caused because these items were the target item on the previous trial, so their activation would have been elevated as a result. Another possible explanation is that because the response times for static subjects remained relatively constant for all non-target responses, these response times reflect processing involved in the mechanisms of the strategy. The updating requirements of the static strategy require less processing following match trials than following non-match trials. But, the processing is the same for all non-match trials, including lures. Thus it is not entirely clear whether response time differences are due to differences in processing requirements between strategies or as a result of the familiarity of the current item.

Finally, evidence was inconclusive as to how stimulus distributions affected the choice of strategies. The interaction of stimulus distribution and trained strategy on response times in experiment 3 showed that static subjects were faster than rolling subjects in skewed sequences but slower in even sequences. Non-trained subjects in experiment 3, however, did not express a preference for the static strategy for skewed sequences, and analysis of performance data for these subjects showed no significant differences related to their choice of strategy.

These results may reflect the generally unreliable nature of self-reports (especially concerning cognitive strategies). Response time data suggest that response times for non-trained subjects matched those of rolling subjects for even sequences, and matched response times of static subjects for skewed sequences. But this data is far from conclusive. Non-trained subjects would have chosen the most comfortable strategy. By contrast, trained subjects were asked to perform a specific strategy, regardless of how natural it was for them. For instance, someone who would have otherwise used the static strategy might perform more poorly when trained to use the rolling strategy than another subject who would have naturally chosen rolling if given the choice.

7.2 Conclusion

This current work has the potential to redefine the constructs thought to be involved in this common paradigm. The n-back is more than a simple measure of working memory capacity. It is a complex test of the ability to match control processes to the demands of the environment and the dynamics of human memory. The results support the hypothesis that sequence manipulations common in the n-back literature can have an effect on the control strategy used to manage the substantial memory demands of the task. Although other researchers have recognized that the n-back involves both a familiarity and a recollection component (Harbison et al., 2011; Juvina & Taatgen, 2007; Szmalec et al., 2011) they have not considered how variations in the base level frequencies of both trial types and stimulus choice may affect the balance of these control components.

A review of the literature reveals a lack of experimental standards in administering this task. As table 2.1 demonstrates, previous researchers have made drastic changes to the trial type and stimulus frequencies that would occur by choosing stimuli in a random fashion. Given the history of the use of statistical manipulation in the cognitive control and memory literature, it seems likely that these manipulations would have predictable effects on n-back performance. Consistent with the general predictions of the DMC theory, random selection of stimuli (with trial type constraints) leads to skewed sequences that force people to change control strategies from primarily proactive to reactive depending on how likely upcoming stimuli are to lead to conflict between competing strategies.

The solution to these issues is not necessarily to always present perfectly counterbalanced stimuli, but rather to conduct an investigation to determine the role these issues play in determining how people approach a task. The overall goal of cognitive research should be to eventually apply findings from the laboratory to the real world. As the real world is not a perfectly uniform place, it is critically important to understand how the mind deals with environmental change. Anderson's rational analysis (Anderson, 1990, 1991) and other research focusing on soft constraints (Fu & Gray, 2006) detail how we use statistical properties of our environment to guide decision-making in the real world. The human memory system, sensitive to recency and frequency effects, allows us to more easily operate in a world where statistical properties of the environment are predictable and stable.

The experiments of this dissertation demonstrated that the ability to match cognition to the nature of the environment translates to performance in the n-back task. The n-back, long thought to be a pure memory task with little to no interactive value, may prove to be just as dependent on the structure of the environment as any of the perceptual/motor interactive tasks used in human factors research. Support for the hypotheses offered here would provide important insight into the interactions between the task environment, memory function, and cognitive control in general. As these factors play into nearly every task we perform, this study is applicable to a broad range of paradigms which as of yet have not received the scrutiny necessary to determine how the task environment affects cognitive performance.

7.3 Recommendations For n-back Researchers

Although the issues raised herein apply to other tasks, it is imperative that researchers using the n-back consider issues related to experimental design more thoroughly than they have in the past. This section will propose recommendations for researchers using n-back going forward.

The results of the current work suggest two issues of concern for researchers. First, common methods of sequence generation results in extremely skewed sequences. The experiments of this dissertation demonstrated that these sequences lead to drastic changes in the use of proactive control. The n-back can be seen as a measure of the ability to sustain proactive control mechanisms over time. Researchers interested in this ability should attempt to avoid the use of skewed sequences. Second, differences in strategy selection (e.g. static vs rolling) corresponded to differences in performance. Again, researchers interested in using the n-back as a measure of control should be careful to control for strategy selection.

The vast majority of n-back researchers have used stimulus selection procedures that guarantee the creation of skewed sequences. It is true that the skewed sequences used in this work were selected from the upper tail of the distribution (e.g. they are slightly more skewed than the average sequence). These sequences were chosen to reduce variability which would naturally occur using common methodologies. That variation is troubling in its own right, as some subjects in past n-back studies surely were presented with sequences more skewed than other subjects. The use of these methods is even more troubling when the length of the sequences is taken into consideration. As figure 2.1 demonstrates, shorter sequences commonly used in the literature create more severely skewed sequences. Thus the following recommendations will help to standardize n-back research and eliminate much of the variability caused by sequence distribution:

7.3.1 Sequence Generation Algorithm

The best way to reduce the effect of skewed sequences is to use an algorithm similar to the one used to create the even sequences in the current work. An implementation of this algorithm can be obtained by contacting the author.

There are potential reasons for using skewed sequences. Researchers interested in studying the ability to recognize statistical regularities should create skewed sequences using the methods described in this work. An algorithm for creating skewed sequences implemented in Python is included in the appendix. Researchers using skewed sequences may also wish to use the correlation between local lumpiness and response time as a measure of this adaptation. The R code used to calculate local lumpiness is also included in the appendix.

7.3.2 Sequence Length

Short sequence lengths can exacerbate issues related to sequence skew. Many previous studies used blocks as short as 20 trials. Figure 2.1 clearly demonstrates that these short sequences usually feature 2 stimuli which are chosen on over 50% of trials. These common stimuli take up the lions share of target trials. A prominent example is Jaeggi's famous dual n-back task (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). The results of this study have been called into question (Redick et al., 2012) on various methodological grounds.

We can add the use of extremely short skewed sequences to the list of methodological complaints. Jaeggi reports some subjects who with adaptive training were able to successfully complete n-back blocks with n as large as 8. Given the skewed nature of her stimulus sequences, it is likely that these subjects improved their performance by adapting control strategies to the nature of the specific task, not by any expansion of working memory capacity.

The problem of skewed sequences can be attenuated by using longer sequences. All three experiments demonstrated that people reach stable performance by the 3rd 64 trial block. Thus I recommend the use of longer blocks of trials. Time constraints should be handled by reducing the number of blocks performed, not by reducing the number of trials per block.

7.3.3 Strategies

The most striking finding was the discovery of the static strategy, which led to superior performance in both experiments 2 and 3, regardless of whether or not subjects were trained to use it. No previous study has controlled for strategy differences in the n-back. Researchers ignore this critical source of variability at their peril. Given the large differences in performance elicited by strategy differences, researchers must make design decisions to address this issue.

First, all n-back studies going forward should use post experiment interviews to determine which strategy each subject actually used. Researchers should report the proportion of subjects who used each strategy and offer data analysis to determine if significant differences in performance correlate with strategy choice.

Second, researchers should decide whether to guide or train subjects to use a specific strategy in accordance with the goals of the study. Researchers not interested in strategy differences may be well served by training all subjects to use the rolling strategy. This will reduce unwanted variability and make data analysis easier to understand.

Training in the rolling strategy may also be useful to researchers interested only using the n-back as a measure of response inhibition. The experiments of the current work demonstrate that the rehearsal process of the rolling strategy cause increased conflict on lure trials that does not seem to occur when using the static strategy.

On the other hand, researchers interested in how people develop and choose strategies may find the n-back to be fertile ground. One of the contributions of this work is to show that n-back strategies are relatively easy to train and cause easily identifiable differences in behavioral data. More research is necessary to determine if these strategy differences may be correlated with differences in fluid intelligence as has been suggested in several previous studies (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Burgess & Braver, 2010; Burgess, Gray, et al., 2011; Gray et al., 2003). It is also unclear if strategy differences could explain why the n-back generally fails to correlate with other complex span memory tasks (Kane et al., 2007) as these other tasks may not be affected by such large strategy differences.

7.4 Future Directions

This work represents a first step in the critical evaluation of the methods and results of previous work on the n-back task. The current work was limited by its reliance on accuracy and response time data. Future research should attempt to replicate these results using other methodologies, such as cognitive modeling and neuro-monitoring (EEG). In addition to replicating these results, future studies should attempt to model the effects discussed in this work and answer questions posed by this work.

7.4.1 Other Methods

The current work revealed several differences in behavioral measures related to stimulus distribution and strategy selection. Replications of this work using EEG or fMRI would provide further evidence as to how these results relate to the cognitive control functions of the brain. Assuming replication, EEG studies could study the change in the n2 waveform of the ERP in relation to lure trials. The n2 is thought to reflect differences in reaction to conflict in conflict monitoring tasks such as th flanker and Stroop tasks (Yeung, Bogacz, et al., 2004; Yeung, Ralph, & Nieuwenhuis, 2007) If differences related to stimulus distribution are indicitave of control differences, then the n2 waveform should vary in correlation with the RT difference results found in the current work. fMRI studies could further support this view by demonstrating changes in activity in dorsolateral prefrontal cortex and anterior cingulate cortex related to stimulus distribution and trained strategies. These studies could help to generalize the results of this work beyond the n-back, showing effects of environmental variation on domain general brain functions.

7.4.2 Cognitive Modeling

The results of this work are sufficiently complex to pose a challenge to current models of the n-back. Cognitive models able to replicate the finding of decreased proactive control for skewed sequences would help to explain why this phenomena was so strong. An attempt at modeling the choice of control strategies in the n-back would also significantly enhance our understanding of control choice in the real world.

The current work provides a road map for measurement of the performance of n-back models. Modelers should attempt to replicate the effects related to control mode and strategy choice. Namely, a successful model should find a positive relationship between the amount of rehearsal and the difference between n-l lure and match response times. Further, attempts to model proactive control in the n-back should find an increase in n+1 lure difference RTs with no significant decrease in RTs for distractor trials as lag increases.

7.4.3 New/Unanswered Questions

An obvious extension of this work is to consider some of the claims that are based primarily upon the results of prior n-back studies. Perhaps the most troubling of these is the argument that n-back training results in an expansion of "working memory capacity" which leads to general increases in cognitive ability and fluid intelligence (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008).

The results of Jaeggi's 2008 study have been questioned and attempted replications have been inconsistent at best (Jaeggi, Buschkuehl, Perrig, & Meier, 2010; Redick et al., 2012). Still, follow up studies have replicated the finding of a training effect in the n-back task in which some subjects are able to accurately perform n-backs at levels of 8-back and above after 20 hours of practice. A long-term training study of the nback using the sequence manipulations proposed here has the potential of determining what aspects of task performance and control are responsible for this increase in performance. It is possible that improvement in n-back performs relates more to the ability to judge relative familiarity and manage conflict than the ability to simply hold multiple items in working memory. Such a result would represent an important contribution as it would emphasize the importance that factors in the environment have on cognitive control and learning.

Additionally, the results of the current work strongly suggest a contribution of strategy development and selection on performance. Thus it is possible that training regimens focusing on developing efficient strategies will outperform those focusing on gains in performance. Tasks beyond the n-back should also be investigated to see how environmental manipulation and strategic choice lead to individual differences in performance. Studies such as these have potential to redefine the constructs we use to measure and assess mental performance and make me excited to spend a career investigating these issues.

References

Abelson, R. P. (2012). *Statistics as principled argument*. New York, NY: Psychology Press.

Altmann, E. M. & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological Review*, *115*(3), 602–39.

Anderson, J. R. (1990). *The adaptive character of thought*. Studies in cognition. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*(3), 471–517.

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.

Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: evidence from positron emission tomography. *Psychological Science*, *7*(1), 25–31.

Baddeley, A. (1984). The fractionation of human memory. *Psychological Medicine*, *14*(2), 259–264.

Baddeley, A. (1992a). Consciousness and working memory. *Consciousness and Cognition*, *1*(1), 3–6.

Baddeley, A. (1992b). Is working memory working?- The 15th Bartlett Lecture. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology*, *44*(1), 1–31.

Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *51*(4), 819–853.

Baddeley, A. & Scott, D. (1971). Short term forgetting in the absence of proactive interference. *Quarterly Journal of Experimental Psychology*, *23*, 275–283.

Bailey, K., West, R., & Anderson, C. a. (2010). A negative association between video game experience and proactive cognitive control. *Psychophysiology*, *47*(1), 34–42.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 105–112.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*, *5*(1), 49–62.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 76–106). New York, NY: Oxford University Press.

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(18), 7351–6.

Brouwer, A.-M., Hogervorst, M. a., van Erp, J. B. F., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering*, *9*(4), 045008.

Burgess, G. C. & Braver, T. S. (2010). Neural mechanisms of interference control in working memory: effects of interference expectancy and fluid intelligence. *PloS one*, *5*(9), e12861.

Burgess, G. C., Gray, J. R., Conway, A. R. a., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology. General*, *140*(4), 674–92.

Chatham, C. H., Herd, S. a., Brant, A. M., Hazy, T. E., Miyake, A., O'Reilly, R., & Friedman, N. P. (2011). From an executive network to executive control: a computational model of the n-back task. *Journal of Cognitive Neuroscience*, *23*(11), 3598–619.

Chen, Y.-N., Mitra, S., & Schlaghecken, F. (2008). Sub-processes of working memory in the n-back task: an investigation using ERPs. *Clinical Neurophysiology : Official journal of the International Federation of Clinical Neurophysiology*, *119*(7), 1546–59.

Crowley, K., Shrager, J., & Siegler, R. S. (1997). Strategy Discovery as a Competitive Negotiation between Metacognitive and Associative Mechanisms. *Developmental Review*, *489*(17), 462–489.

Fu, W.-T. & Gray, W. D. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology*, *52*(3), 195–242.

Gigerenzer, G. & Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–143.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316–22.

Harbison, J. I., Atkins, S. M., & Dougherty, M. R. (2011). N-back training task performance : analysis and model experiment : training data. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 120–125). Austin, TX: Cognitive Science Society.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: making working memory work. *Neuroscience*, *139*(1), 105–18.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2010). Neural mechanisms of acquired phasic dopamine responses in learning. *Neuroscience and Biobehavioral Reviews*, *34*(5), 701–20.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the n-back task as a working memory measure. *Memory*, *18*(4), 394–412.

Juvina, I. & Taatgen, N. A. (2007). Modeling control strategies in the n-back task. In R. Lewis, T. Polk, & J. Laird (Eds.), *Proceedings of the 8th International Conference on Cognitive Modeling* (pp. 73–78). Oxford, UK: Taylor & Francis/Psychology Press.

Kane, M. J., Conway, A. R. a., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(3), 615–22.

Lawrence, M. A. (2011). *Ez: easy analysis and visualization of factorial experiments.* R package version 3.0-0, Retrieved from http://CRAN.R-project.org/package=ez, Last accessed on 2014-10-2.

Lewis, R. L., Shvartsman, M., & Singh, S. (2013). The adaptive nature of eye movements in linguistic tasks: how payoff and architecture shape speed-accuracy trade-offs. *Cognitive Science*, *5*(3), 581–610.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *27*(3), 817–35.

Miyake, A. & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenence and executive control.* New York, NY: Cambridge University Press.

Nairne, J. S. (2002). Remembering over the short-term : the case against the standard model. *Annual Review of Psychology*, *53*, 53–81.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59.

Oztekin, I., Davachi, L., & McElree, B. (2010). Are representations in working memory distinct from representations in long-term memory? neural evidence in support of a single store. *Psychological Science*, *21*(8), 1123–33.

R Core Team. (2013). *R: a language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org/ Last accessed 2014-10-2.

Reder, L. M., Weber, K., Shang, J., & Vanyukov, P. M. (2003). The adaptive character of the attentional system : statistical sensitivity in a target localization task. *Journal of Experimental Psychology*, *29*(3), 631–649.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., … Engle, R. W. (2012). No Evidence of Intelligence Improvement After Working Memory Training: A Randomized, Placebo-Controlled Study. *Journal of Experimental Psychology, General*, *142*(2), 359–379.

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT press.

Speer, N. K., Jacoby, L. L., & Braver, T. S. (2003). Strategy-dependent changes in memory: effects on behavior and brain activity. *Cognitive, Affective & Behavioral Neuroscience*, *3*(3), 155–67.

Szmalec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(1), 137–51.

Yeung, N., Bogacz, R., Holroyd, C. B., & Cohen, J. D. (2004). Detection of synchronized oscillations in the electroencephalogram: An evaluation of methods. *Psychophysiology*, *41*(6), 822–832.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection : conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959.

Yeung, N., Ralph, J., & Nieuwenhuis, S. (2007). Drink alcohol and dim the lights: the impact of cognitive deficits on medial frontal cortex function. *Cognitive, Affective & Behavioral Neuroscience*, *7*(4), 347–55.

Appendices

Appendix A
Sequences

A.1 Sample Sequences

The following sequence are samples of those used in the experiments. The sequences list the letter presented, the trial type it represents and it's lag position

Sample Even - 24 Sequence

Trial	Letter	Type
1	R	Dist
2	B	Dist
3	Z	Dist
4	R	Target
5	B	Target
6	F	Dist
7	J	Dist
8	R	Lure4
9	F	Target
10	R	Lure2
11	M	Dist
12	M	Dist
13	F	Lure4
14	X	Dist
15	J	Dist
16	Z	Dist
17	X	Target
18	J	Target
19	B	Dist
20	X	Target
21	B	Lure2
22	F	Dist
23	X	Target

Trial	Letter	Type
24	M	Dist
25	H	Dist
26	Z	Dist
27	M	Target
28	X	Dist
29	Z	Target
30	M	Target
31	X	Target
32	B	Dist
33	M	Target
34	B	Lure2
35	H	Dist
36	X	Dist
37	H	Lure2
38	R	Dist
39	Z	Dist
40	J	Dist
41	X	Dist
42	Z	Target
43	M	Dist
44	R	Dist
45	Z	Target

Trial	Letter	Type
46	H	Dist
47	J	Dist
48	Z	Target
49	R	Dist
50	J	Target
51	Z	Target
52	R	Target
53	J	Target
54	F	Dist
55	B	Dist
56	M	Dist
57	H	Dist
58	B	Target
59	F	Dist
60	H	Target
61	B	Target
62	F	Target
63	H	Target
64	R	Dist
65	J	Dist
66	F	Lure4
67	H	Lure4

Sample Even - 8 Sequence

Trial	Letter	Type
1	X	Dist
2	F	Dist
3	M	Dist
4	X	Target
5	M	Lure2
6	Z	Dist
7	Z	Dist
8	X	Lure4
9	B	Dist
10	X	Lure2
11	J	Dist
12	Z	Dist
13	F	Dist
14	R	Dist
15	H	Dist
16	R	Lure2
17	X	Dist
18	H	Target
19	Z	Dist
20	R	Lure4
21	M	Dist
22	F	Dist
23	Z	Lure4

Trial	Letter	Type
24	H	Dist
25	J	Dist
26	X	Dist
27	J	Lure2
28	H	Lure4
29	Z	Dist
30	H	Lure2
31	F	Dist
32	Z	Target
33	F	Lure2
34	Z	Lure2
35	B	Dist
36	M	Dist
37	F	Lure4
38	M	Lure2
39	F	Lure2
40	H	Dist
41	X	Dist
42	H	Lure2
43	B	Dist
44	X	Target
45	R	Dist

Trial	Letter	Type
46	B	Target
47	J	Dist
48	B	Lure2
49	R	Lure4
50	M	Dist
51	J	Lure4
52	H	Dist
53	M	Target
54	R	Dist
55	J	Lure4
56	X	Dist
57	M	Lure4
58	B	Dist
59	M	Lure2
60	F	Dist
61	J	Dist
62	B	Lure4
63	F	Target
64	R	Dist
65	J	Lure4
66	B	Lure4
67	R	Target

Sample Skewed - 24 Sequence

Trial	Letter	Type
1	F	Dist
2	X	Dist
3	Z	Dist
4	R	Dist
5	X	Target
6	M	Target
7	R	Target
8	X	Target
9	M	Target
10	Z	Dist
11	X	Target
12	Z	Lure2
13	Z	Target
14	R	Dist
15	Z	Lure2
16	M	Dist
17	B	Dist
18	Z	Target
19	H	Target
20	Z	Lure2
21	R	Dist
22	R	Dist
23	X	Dist
24	R	Lure2
25	Z	Dist
26	X	Target
27	R	Target
28	X	Lure2
29	X	Target
30	R	Target
31	F	Dist
32	M	Dist
33	R	Target
34	Z	Dist
35	F	Lure4
36	M	Lure4
37	Z	Target
38	F	Target
39	M	Target
40	M	Lure4
41	M	Lure2
42	F	Lure4
43	B	Dist
44	X	Dist
45	J	Dist
46	B	Target
47	Z	Dist
48	J	Target
49	M	Dist
50	Z	Target
51	H	Dist
52	H	Dist
53	Z	Target
54	M	Dist
55	B	Dist
56	R	Dist
57	J	Dist
58	B	Target
59	H	Dist
60	M	Dist
61	R	Dist
62	J	Dist
63	H	Lure4
64	R	Target
65	J	Target
66	Z	Dist
67	M	Dist

Sample Skewed-8 Sequence

Trial	Letter	Type
1	X	Dist
2	F	Dist
3	R	Dist
4	H	Dist
5	X	Lure4
6	Z	Dist
7	X	Lure2
8	Z	Lure2
9	J	Dist
10	H	Dist
11	X	Lure4
12	Z	Lure4
13	J	Lure4
14	X	Target
15	X	Lure4
16	Z	Lure4
17	R	Dist
18	X	Target
19	X	Lure4
20	M	Dist
21	M	Dist
22	X	Target
23	M	Lure2

Trial	Letter	Type
24	F	Dist
25	B	Dist
26	X	Lure4
27	X	Dist
28	R	Dist
29	Z	Dist
30	M	Dist
31	M	Dist
32	Z	Target
33	M	Lure2
34	Z	Lure2
35	X	Dist
36	H	Dist
37	Z	Target
38	F	Dist
39	X	Lure4
40	X	Dist
41	Z	Lure4
42	M	Dist
43	M	Dist
44	M	Lure2
45	Z	Lure4

Trial	Letter	Type
46	Z	Dist
47	X	Dist
48	B	Dist
49	X	Lure2
50	F	Dist
51	R	Dist
52	F	Lure2
53	R	Lure2
54	F	Lure2
55	Z	Dist
56	F	Lure2
57	B	Dist
58	J	Dist
59	F	Target
60	R	Dist
61	B	Lure4
62	R	Lure2
63	X	Dist
64	X	Dist
65	M	Dist
66	J	Dist
67	B	Dist

A.2 Skewed Sequence Code

```python
import time, random
from collections import deque

class NBack(object):
    """N-Back Task"""

    ALPHA1 = 1
    ALPHA2 = 2

    def __init__(self, n=3, pT=24, pL1=4, pL2=4, pD=32, trials=64, type=ALPHA1):
        """Initialize an NBack object; takes 3 optional arguments:
        n       The set size of stimlui to remember
        pT      The probability that the n-th back will match
    pL1     The probability that the n-1-th back will match
    pL2     The probability that the n+1-th back will match
    type    The stimuli pool to use"""

        super(NBack, self).__init__()

        self.pT = pT
        self.pD=pD
        self.pL1 = pL1
        self.pL2 = pL2
        self.n = n
        self.trials = trials
        self.t_rem = trials
        self.pool = self._generate_stimuli_pool(type)
        self.buffer = self._init_buffer()

    def _init_buffer(self):
      buf = deque()

      for i in range(1,5):
        item = random.choice(self.pool)

```

```
      buf.appendleft(item)

    return buf

  def _generate_stimuli_pool(self, type):

      pool = None

      if type == self.ALPHA1:
        pool = ['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H']

      elif type == self.ALPHA2:
        pool = ['b', 'c', 'd', 'g']

      return pool

  def _next_TT(self):
      """Returns the next trial type"""

      num = random.randrange(1,self.t_rem + 1)
      L1_lim = self.pL1
      L2_lim = L1_lim + self.pL2
      T_lim = L2_lim + self.pT

      if num <= L1_lim:
        """Choose n-1 lure"""
        TT='L1'
      elif num > L1_lim and num <= L2_lim:
        """Choose n+1 lure"""
        TT='L2'
      elif num > L2_lim and num <= T_lim:
        """Choose Target"""
        TT='T'
      else:
        TT='D'

      return(TT)

```

```
    def check_TT(self,trial_type):
      """Check if Trial Type is OK"""

      trial_OK = False

      if trial_type == 'T' and self.pT > 0:
        trial_OK = True
        self.pT-=1
        self.t_rem-=1

      elif trial_type == 'L1' and self.pL1 > 0:

        if self.buffer[1] != self.buffer[2]:

          trial_OK = True
          self.pL1-=1
          self.t_rem-=1

        elif self.pT==0 and self.pD==0 and self.pL2==0:
          trial_OK = True
          self.pL1-=1
          self.t_rem-=1

      elif trial_type == 'L2' and self.pL2 > 0:
        if self.buffer[3] != self.buffer[2] and self.buffer[3] != self.
buffer[1]:

          trial_OK = True
          self.pL2-=1
          self.t_rem-=1

        elif self.pT==0 and self.pD==0 and self.pL1==0:
          trial_OK = True
          self.pL2-=1
          self.t_rem-=1

      elif trial_type =='D':
        trial_OK = True
```

```
      self.pD-=1
      self.t_rem-=1

    return(trial_OK)

  def next(self):

    Trial_Type = self._next_TT()
    countfalse=0

    while (self.check_TT(Trial_Type) == False and countfalse<50):
      Trial_Type = self._next_TT()
      print("%s %s /n" % (Trial_Type,countfalse))
      countfalse +=1
      if countfalse > 48:
        Trial_Type = 'D'

    if Trial_Type == 'T':
      new = self.buffer[2]
    elif  Trial_Type == 'L1':
      new = self.buffer[1]
    elif Trial_Type == 'L2':
      new = self.buffer[3]
    elif Trial_Type == 'D':
      new = random.choice(self.pool)
      check_list = [self.buffer[1],self.buffer[2],self.buffer[3]]

      while new in check_list:
        new = random.choice(self.pool)

    self.buffer.appendleft(new)
    value = (new, Trial_Type, self.t_rem)
    s = str(value)

    return new


  def respond(self, ans):
```

```
        """Pass the ans argument to indicate if n-th back was a match.
        Returns None if there are not yet n+1 items in the buffer else it
        returns the correctness of ans."""

        ret = None
        if len(self.buffer) == self.n + 1:
            match = self.buffer[0] == self.buffer[1]
            if match == ans:
                ret = True
            else:
                ret = False
        return ret

if __name__ == "__main__":

  f=open('/Users/jasonralphold/Documents/skewed_24', 'w')
  j=0
  while j < 20000:
    exp = NBack()
    while exp.t_rem > 1:
      x=exp.next()
    f.write("%s \n" % exp.buffer)
    print(j)

    j+=1
```

Appendix B
R Code

This sections includes the R code used to analyze data in this dissertation

B.0.1 Lumpiness Code

Functions for calculating local lumpiness used for all three experiments

```r
# Takes a vector and returns a matrix with n rows filled with the previous n items

prevrows <- function(data,n)
{sapply(1:n,function(x) c(rep(NA,x),head(data,-x)))}

# Takes a vector and returns a list of the counts of items in each row of 8 previous items

stim_counts <- function(data,n)
{apply(prevrows(data,n),1,function(x) tapply(x,x,length))}

# Takes a vector and returns a lumpiness calculation for the previous n items of each position in the vector

lumps <- function(data,n)
{sapply(stim_counts(data,n),function(x) sum(x-1))}
```

B.0.2 ANOVA

The signal detection ANOVAs were performed by using the native aov function because this allowed for a trend analysis to determine if stable performance was reached after 3 blocks. An example of these analyses is below:

```r
e1_dp_aov <- aov(dp ~ (ratio*dist*block) + Error(subj/block));
```

```
summary(e1_dp_aov, split = list(BLK = list(linear = 1, quadratic
    = 2, cubic = 3))));
```

All other analyses were performed by using the ezANOVA function of the ez package. An example is below:

```
e1_diff <- ezANOVA(data, dv = response_time, wid = subject,
    between = .(ratio, distribution), within = .(lure_position),
    type=3)
```

All ezANOVA analyses were performed with type 3 sum of squares. η_p^2 values were calculated manually using the output from the ezANOVA functions.

Appendix C
Experiment 1

C.1 Data and Statistics

Table C.1: Summary Table (mean and sd) for SDT analyses (d' and c)

		Lure Heavy		Target Heavy	
	Stimulus Distribution	mean	sd	mean	sd
d'	Even	2.24	1.32	2.80	1.23
	Skewed	2.69	1.51	3.24	0.90
crit	Even	0.63	0.58	0.15	0.29
	Skewed	0.48	0.58	0.20	0.32

Table C.2: Experiment 1 ANOVA table for sensitivity (d') with linear trend analysis for block

Error - Subject	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	55.42	55.42	4.81	0.0317	0.06
Stimulus Distribution	1	43.86	43.86	3.80	0.0551	0.05
T:L Ratio:Stimulus Distribution	1	0.12	0.12	0.01	0.9193	
Residuals	70	807.15	11.53			

Error - Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	11	68.58	6.23	9.60	0.0000	0.12
linear	1	51.38	51.38	79.13	0.0000	
quadratic	1	11.16	11.16	17.18	0.0000	
cubic	1	2.93	2.93	4.52	0.0339	
T:L Ratio:Block	11	3.58	0.33	0.50	0.9030	
linear	1	0.41	0.41	0.64	0.4255	
quadratic	1	0.38	0.38	0.59	0.4416	
cubic	1	1.13	1.13	1.74	0.1874	
Stimulus Distribution:Block	11	12.96	1.18	1.81	0.0479	0.03
linear	1	2.39	2.39	3.68	0.0555	
quadratic	1	0.86	0.86	1.32	0.2506	
cubic	1	1.77	1.77	2.72	0.0993	
T:L Ratio:Distribution:Block	11	6.80	0.62	0.95	0.4887	
linear	1	0.18	0.18	0.27	0.6006	
quadratic	1	0.84	0.84	1.29	0.2570	
cubic	1	0.07	0.07	0.11	0.7370	
Residuals	770	500.00	0.65			

Table C.3: Experiment 1 - ANOVA table - d' for blocks 4-12 (after eliminating first 3 blocks)

Error - Subject	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	48.00	48.00	5.04	0.0280	0.07
Stimulus Distribution	1	33.01	33.01	3.46	0.0669	0.05
T:L Ratio:Stimulus Distribution	1	0.01	0.01	0.00	0.9791	
Residuals	70	666.85	9.53			

Error - Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	8	7.10	0.89	1.38	0.2039	
T:L Ratio:Block	8	1.35	0.17	0.26	0.9776	
Stimulus Distribution:Block	8	12.84	1.60	2.49	0.0117	0.03
T:L Ratio:Distribution:Block	8	5.37	0.67	1.04	0.4041	
Residuals1	560	361.30	0.65			

Table C.4: Experiment 1- ANOVA table for criterion (c) (all 12 blocks with trend analysis)

Error - Subject	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	29.30	29.30	26.06	0.0000	0.27
Stimulus Distribution	1	0.42	0.42	0.37	0.5448	
T:L Ratio:Distribution	1	1.08	1.08	0.96	0.3303	
Residuals	70	78.69	1.12			

Error - Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	11	4.41	0.40	3.27	0.0002	0.04
linear	1	2.90	2.90	23.69	0.0000	
quadratic	1	0.65	0.65	5.30	0.0215	
cubic	1	0.20	0.20	1.66	0.1978	
T:L Ratio:Block	11	0.88	0.08	0.65	0.7827	
linear	1	0.06	0.06	0.51	0.4755	
quadratic	1	0.01	0.01	0.05	0.8312	
cubic	1	0.54	0.54	4.41	0.0361	
Stimulus Distribution:Block	11	1.29	0.12	0.96	0.4824	
linear	1	0.10	0.10	0.84	0.3595	
quadratic	1	0.07	0.07	0.60	0.4379	
cubic	1	0.12	0.12	0.98	0.3215	
T:L Ratio:Stimulus Distribution:Block	11	3.00	0.27	2.23	0.0117	0.03
linear	1	0.04	0.04	0.32	0.5696	
quadratic	1	0.82	0.82	6.73	0.0097	
cubic	1	0.08	0.08	0.64	0.4245	
Residuals	770	94.38	0.12			

Table C.5: Experiment 1 ANOVA table for criterion analysis (Blocks 4-12 with trend analysis).

Error - Subject	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	22.90	22.90	23.16	0.00	0.25
Stimulus Distribution	1	0.38	0.38	0.38	0.5388	
T:L Ratio:Stimulus Distribution	1	1.46	1.46	1.48	0.2284	
Residuals	70	69.24	0.99			

Error - Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	8	2.02	0.25	2.13	0.0316	0.03
linear	1	0.04	0.04	0.37	0.5439	
quadratic	1	0.06	0.06	0.53	0.4685	
cubic	1	0.02	0.02	0.15	0.7032	
T:L Ratio:Block	8	0.49	0.06	0.52	0.8424	
linear	1	0.05	0.05	0.44	0.5070	
quadratic	1	0.00	0.00	0.02	0.8860	
cubic	1	0.01	0.01	0.07	0.7988	
Stimulus Distribution:Block	8	1.23	0.15	1.29	0.2444	
linear	1	0.43	0.43	3.58	0.0589	
quadratic	1	0.13	0.13	1.11	0.2927	
cubic	1	0.16	0.16	1.33	0.2488	
T:L Ratio:Stimulus Distribution:Block	8	2.52	0.31	2.65	0.0074	0.04
linear	1	0.34	0.34	2.88	0.0904	
quadratic	1	0.63	0.63	5.27	0.0221	
cubic	1	0.05	0.05	0.38	0.5368	
Residuals	560	66.55	0.12			

Table C.6: Summary Table for Match-Lure response times.

		T:L Ratio			
		Lure Heavy		Target Heavy	
Lure Type	Stimulus Distribution	mean	sd	mean	sd
N-1	Even	37.21	127.22	102.68	82.95
	Skewed	– 14.94	102.12	35.78	54.38
N+1	Even	– 47.82	122.30	80.57	103.94
	Skewed	– 65.08	84.17	88.33	76.47

Table C.7: ANOVA table for analysis of response time differences between lure and match trials

Error - Subject	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	373757.13	373757.13	22.67	0.0000	0.24
Stimulus Distribution	1	38056.31	38056.31	2.31	0.1332	
T:L Ratio:Stimulus Distribution	1	242.64	242.64	0.01	0.9038	0
Residuals	70	1153875.92	16483.94			

Error - Subject:Lure Position	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Lure Position	1	23075.09	23075.09	11.85	0.0010	0.14
T:L Ratio:Lure Position	1	62799.32	62799.32	32.26	0.0000	0.31
Stimulus Distribution:Lure Position	1	27701.92	27701.92	14.23	0.0003	0.16
T:L Ratio:Stimulus Distribution:Lure Position	1	3643.54	3643.54	1.87	0.1757	
Residuals	70	136266.21	1946.66			

Table C.8: Analysis of covariance table for n-1 lure response times by lumpiness

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	30459.68	30459.68	1.64	0.2017	0.004
Stimulus Distribution	1	356193.93	356193.93	19.13	0	0.05
Lumpiness	1	1561.79	1561.79	0.08	0.7723	
T:L Ratio:Stimulus Distribution	1	14633.45	14633.45	0.79	0.3759	
T:L Ratio:Lumpiness	1	716814.09	716814.09	38.49	0.0000	0.09
Stimulus Distribution:Lumpiness	1	9913.74	9913.74	0.53	0.4661	
T:L Ratio:Stimulus Distribution:Lumpiness	1	280179.97	280179.97	15.05	0.0001	0.04
Residuals	367	6834079.13	18621.47			

Appendix D
Experiment 2

D.1 Instructions

3-Back Experiment

Thank You for Paticipating

The Task

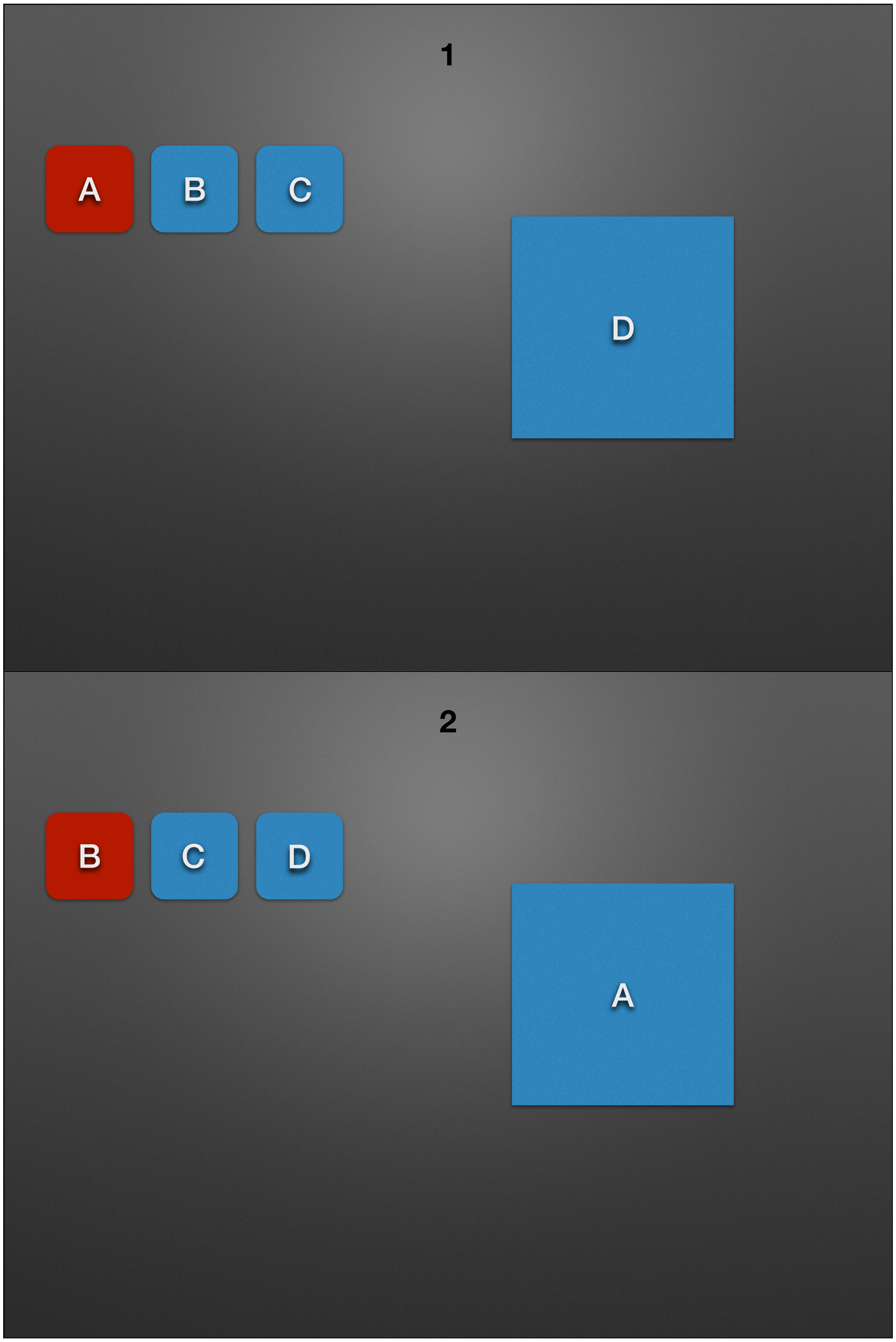
- You will see a sequence of letters (one at a time)
- Determine if each letter is the same or different as the third previous letter.
- If it is the same, press the letter “A”, if different, press “L.”
- For instance, if the sequence is C B A **C B** C, the bolded C and B are matches, and all the others are non-matches.

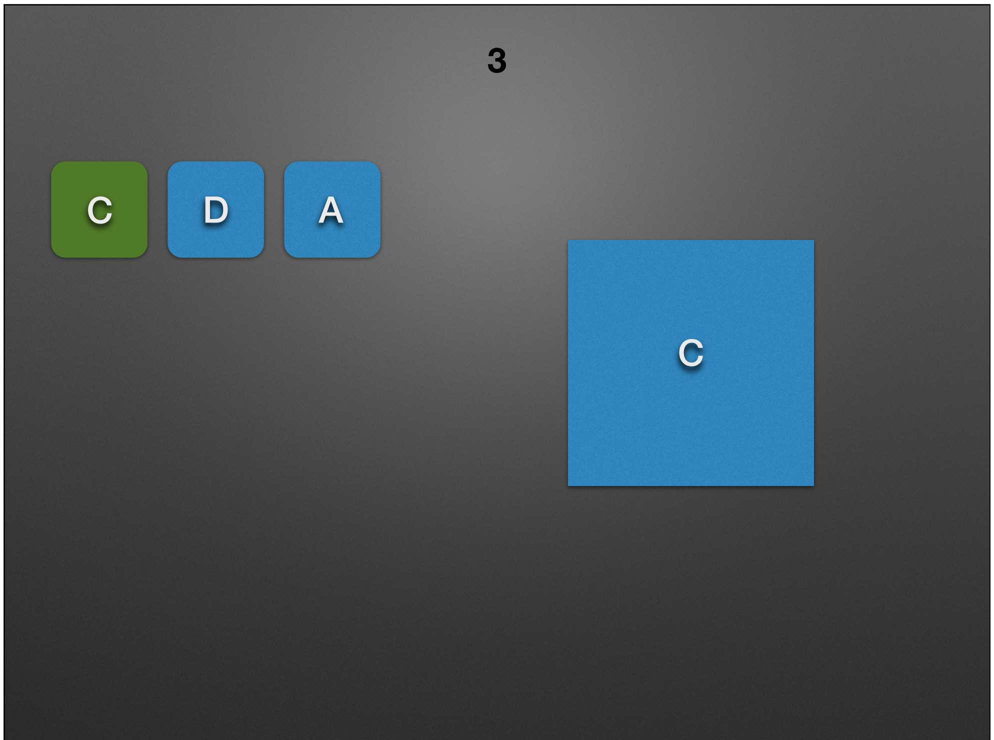
Strategies

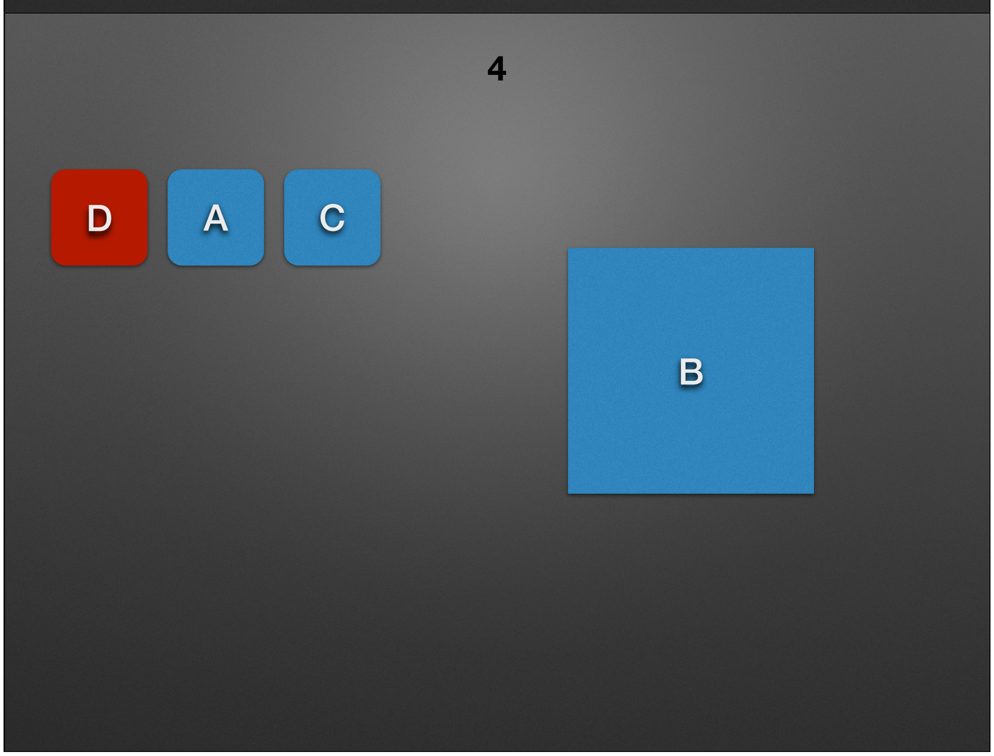
- Your researcher will instruct you in the use of a specific strategy to use
- Please try to use this strategy even if you think there might be a better way to perform the task.
- Try to respond as quickly and accurately as possible, but do not worry if you make mistakes. The most important part of this experiment is to try to stick with the instructed strategy

Rolling Rehearsal Strategy

- Rehearse the previous three letters in sequence (A B C, A B C)
- When a new letter appears, compare it to the first letter in the sequence, then drop the first letter and add the new letter to the end (like a conveyor belt)
  - A B C … B C D … C D B
- During practice, rehearse out loud so the experimenter can tell that you are learning this strategy


1
A
B
C
D
2
B
C
D
A


3
C
D
A
C


4
D
A
C
B

D.2 Data and Statistics

Table D.1: Summary Tables for d' and c

		d'		C	
Training	T:L Ratio	mean	sd	mean	sd
no-training	Target Heavy	2.584	1.072	0.19896	0.2781
	Lure Heavy	2.315	1.366	0.07963	0.4180
rolling	Target Heavy	2.781	1.262	0.11051	0.3052
	Lure Heavy	2.993	1.231	0.10207	0.4045

Table D.2: ANOVA for d' for the full dataset (blocks 1 - 12).

Error - Subject	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	0.97	0.97	0.08	0.7797	
Training	1	33.76	33.76	2.75	0.1026	
T:L Ratio:Training	1	10.30	10.30	0.84	0.3632	
Residuals	56	686.49	12.26			

Error - Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	11	20.70	1.88	3.11	0.0004	0.05
linear	1	5.39	5.39	8.92	0.0029	
quadratic	1	9.01	9.01	14.90	0.0001	
cubic	1	0.82	0.82	1.36	0.2434	
T:L Ratio:Block	11	6.62	0.60	0.99	0.4495	
linear	1	0.51	0.51	0.84	0.3588	
quadratic	1	0.72	0.72	1.18	0.2768	
cubic	1	0.45	0.45	0.74	0.3889	
Training:Block	11	12.09	1.10	1.82	0.0477	0.03
linear	1	1.03	1.03	1.71	0.1915	
quadratic	1	3.68	3.68	6.09	0.0139	
cubic	1	0.32	0.32	0.53	0.4651	
T:L Ratio:Training:Block	11	6.27	0.57	0.94	0.4989	
linear	1	0.46	0.46	0.77	0.3821	
quadratic	1	0.36	0.36	0.60	0.4381	
cubic	1	0.38	0.38	0.63	0.4293	
Residuals	616	372.42	0.60			

Table D.3: ANOVA for d' for the truncated dataset (blocks 4 - 12).

Error - Subject	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	0.16	0.16	0.02	0.8989	
Training	1	33.29	33.29	3.31	0.0742	0.06
T:L Ratio:Training	1	7.06	7.06	0.70	0.4058	
Residuals	56	563.40	10.06			

Error - Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	8	4.46	0.56	0.96	0.4654	
T:L Ratio:Block	8	5.18	0.65	1.12	0.3509	
Training:Block	8	8.11	1.01	1.75	0.0857	0.03
T:L Ratio:Training:Block	8	5.04	0.63	1.09	0.3712	
Residuals	448	259.91	0.58			

Table D.4: Experiment 2 - ANOVA table for criterion - full dataset with trend analysis

Error - Subject	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	0.77	0.77	2.60	0.1124	
Training	1	0.20	0.20	0.68	0.4141	
T:L Ratio:Training	1	0.55	0.55	1.85	0.1787	
Residuals	56	16.56	0.30			

Error - Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	11	3.61	0.33	3.08	0.0005	0.05
linear	1	1.55	1.55	14.55	0.0002	
quadratic	1	0.10	0.10	0.89	0.3450	
cubic	1	0.53	0.53	4.96	0.0263	
T:L Ratio:Block	11	1.91	0.17	1.63	0.0854	0.03
linear	1	0.64	0.64	5.98	0.0147	
quadratic	1	0.01	0.01	0.07	0.7929	
cubic	1	0.02	0.02	0.16	0.6895	
Training:Block	11	1.87	0.17	1.60	0.0948	0.03
linear	1	0.07	0.07	0.70	0.4042	
quadratic	1	0.53	0.53	5.01	0.0256	
cubic	1	0.08	0.08	0.72	0.3970	
T:L Ratio:Training:Block	11	1.86	0.17	1.59	0.0969	0.03
linear	1	0.24	0.24	2.27	0.1322	
quadratic	1	0.04	0.04	0.42	0.5185	
cubic	1	0.11	0.11	1.00	0.3175	
Residuals	616	65.52	0.11			

Table D.5: Experiment 2 - ANOVA table for criterion - truncated dataset with trend analysis

Error - Subject	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	0.14	0.14	0.57	0.4549	
Training	1	0.32	0.32	1.32	0.2556	
T:L Ratio:Training	1	0.68	0.68	2.83	0.0978	0.05
Residuals	56	13.45	0.24			

Error - Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	8	1.90	0.24	2.20	0.0267	0.04
linear	1	0.40	0.40	3.66	0.0565	
quadratic	1	0.04	0.04	0.33	0.5649	
cubic	1	0.10	0.10	0.90	0.3437	
T:L Ratio:Block	8	1.00	0.13	1.16	0.3220	
linear	1	0.15	0.15	1.38	0.2406	
quadratic	1	0.46	0.46	4.21	0.0408	
cubic	1	0.23	0.23	2.14	0.1440	
Training:Block	8	1.03	0.13	1.19	0.3014	
linear	1	0.11	0.11	0.99	0.3199	
quadratic	1	0.08	0.08	0.71	0.3997	
cubic	1	0.00	0.00	0.03	0.8518	
T:L Ratio:Training:Block	8	1.72	0.22	1.99	0.0460	0.03
linear	1	0.78	0.78	7.25	0.0073	
quadratic	1	0.07	0.07	0.61	0.4335	
cubic	1	0.02	0.02	0.18	0.6700	
Residuals	448	48.48	0.11			

Table D.6: Experiment 2 - Summary Table - RT by trial type (lure - match differences)

		N-1 Lure Difference		N+1 Lure Difference	
Training	T:L Ratio	mean	sd	mean	sd
no-training	Target Heavy	141.67	68.20	90.54	82.19
	Lure Heavy	73.03	94.56	24.30	86.98
rolling	Target Heavy	202.29	76.07	103.54	74.74
	Lure Heavy	113.76	64.99	37.31	73.03

Table D.7: ANOVA table for the difference between match and lure trials

Error - Subject	Df	Sum Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	165194.9	14.998	.0028	0.22
Training	1	30228.4	2.900	.094	0.05
T:L Ratio:Training	1	734.9	0.071	.79	
Residuals	56	582140.1			

Error - Subject:Lure Position	Df	Sum Sq	F value	Pr(>F)	η_p^2
Lure Position	1	138062.26	68.13	0.00	0.54
T:L Ratio:Lure Position	1	1512.95	0.55	.46	
Training:Lure Position	1	10598.91	5.11	.028	0.08
T:L Ratio:Training:Lure Position	1	735.88	0.357	.55	
Residuals	56	115551.86			

Table D.8: Experiment 2 - ANCOVA table for n-1 lure trials by lumpiness

Effect	Df	Sum Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	424853.02	17.62	0.0000	0.06
Training	1	5456.28	0.23	0.6346	
Lumpiness	1	29936.72	1.24	0.2661	
T:L Ratio:Training	1	182566.78	7.57	0.0063	0.03
T:L Ratio:Lumpiness	1	25294.79	1.05	0.3066	
Training:Lumpiness	1	12738.36	0.53	0.4679	
T:L Ratio:Training:Lumpiness	1	3034.57	0.13	0.7230	
Residuals	283	6823300.08			

Table D.9: ANOVA table for sensitivity(d') and criterion (c) by selected strategy for non-trained subjects

Sensitivity (d')

Error - Subject	Df	Sum Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	9.62	0.40	0.53	
Strategy	1	62.18	8.95	0.007	0.29
T:L Ratio:Strategy	1	1.91	0.27	0.61	
Residuals	21	149.43			

Error - Subject:Block	Df	Sum Sq	F value	Pr(>F)	η_p^2
Block	8	3.81	0.62	0.76	
T:L Ratio:Block	8	4.13	0.95	0.47	
Strategy:Block	8	3.85	0.68	0.71	
T:L Ratio:Strategy:Block	8	5.26	1.03	0.42	
Residuals	168	107.72			

Criterion (c)

Error - Subject	Df	Sum Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	0.16	0.70	0.41	
Strategy	1	0.02	0.09	0.77	
T:L Ratio:Strategy	1	0.03	0.10	0.76	
Residuals	21	5.91			

Error - Subject:Block	Df	Sum Sq	F value	Pr(>F)	η_p^2
Block	8	0.66	1.17	0.32	
T:L Ratio:Block	8	0.81	0.95	0.47	
Strategy:Block	8	0.61	0.90	0.52	
T:L Ratio:Strategy:Block	8	0.40	0.55	0.82	
Residuals	168	15.34			

Table D.10: ANOVA table for RT differences between lure and match trials for non-trained subjects

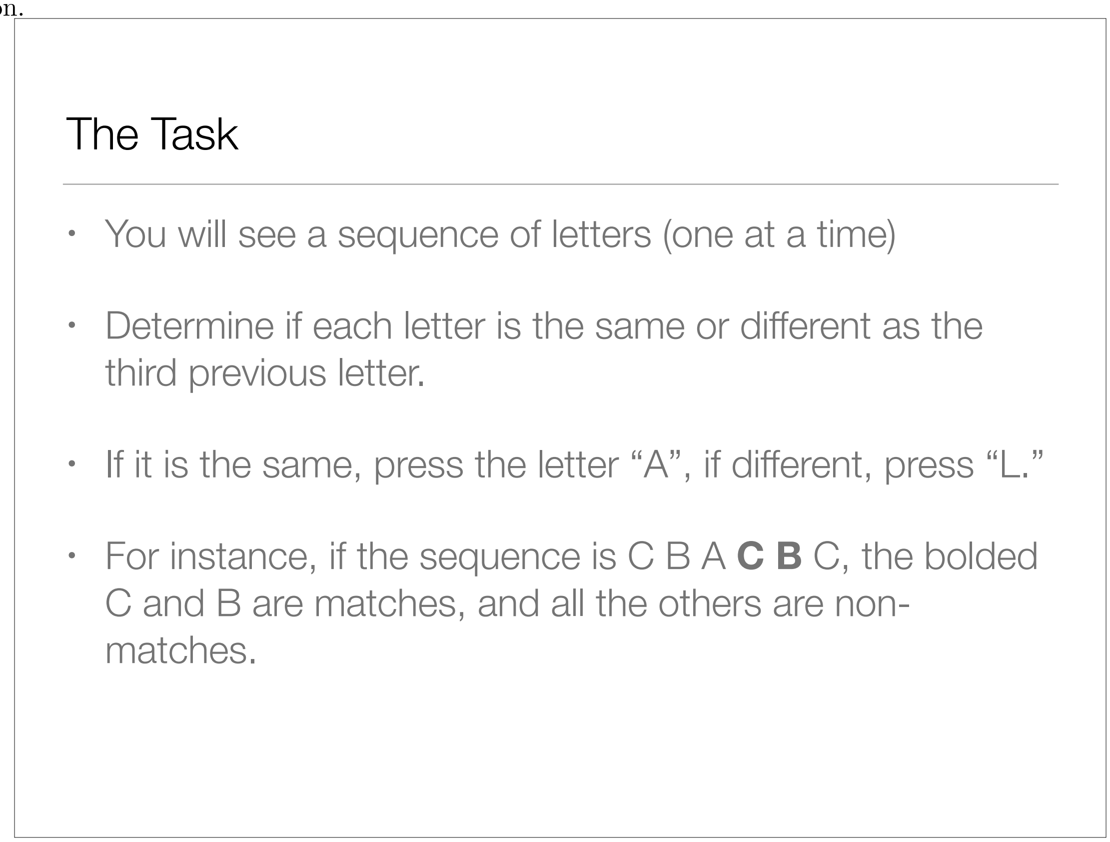
Error - Subject	Df	Sum Sq	F value	Pr(>F)	η_p^2
T:L Ratio	1	94187.57	8.83	0.007	0.27
Strategy	1	199.51	0.062	0.77	
T:L Ratio:Strategy	1	19820.23	1.61	0.22	
Residuals	21	258191.24			

Error - Subject:Lure Position	Df	Sum Sq	F value	Pr(>F)	η_p^2
Lure Position	1	26731.40	11.72	0.0026	0.40
T:L Ratio:Lure Position	1	2249.37	0.68	0.42	
Strategy:Lure Position	1	1598.44	0.94	0.34	
T:L Ratio:Strategy:Lure Position	1	561.54	0.30	0.59	
Residuals	21	39376.50			

Appendix E
Experiment 3

E.1 Instructions

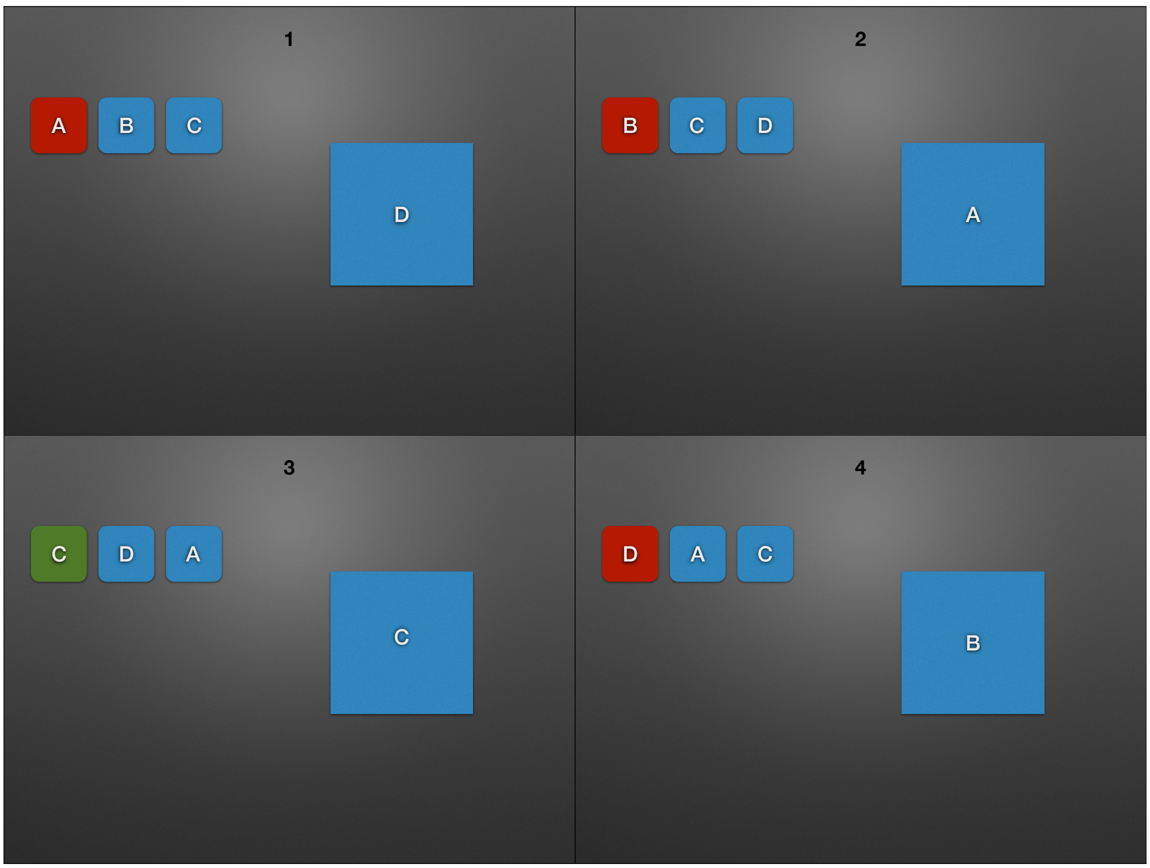
These slides are the generic instructions shown to all subjects in experiment 3. After these slides subjects were instructed using the strategy specific schematic diagrams depicted in the next section.



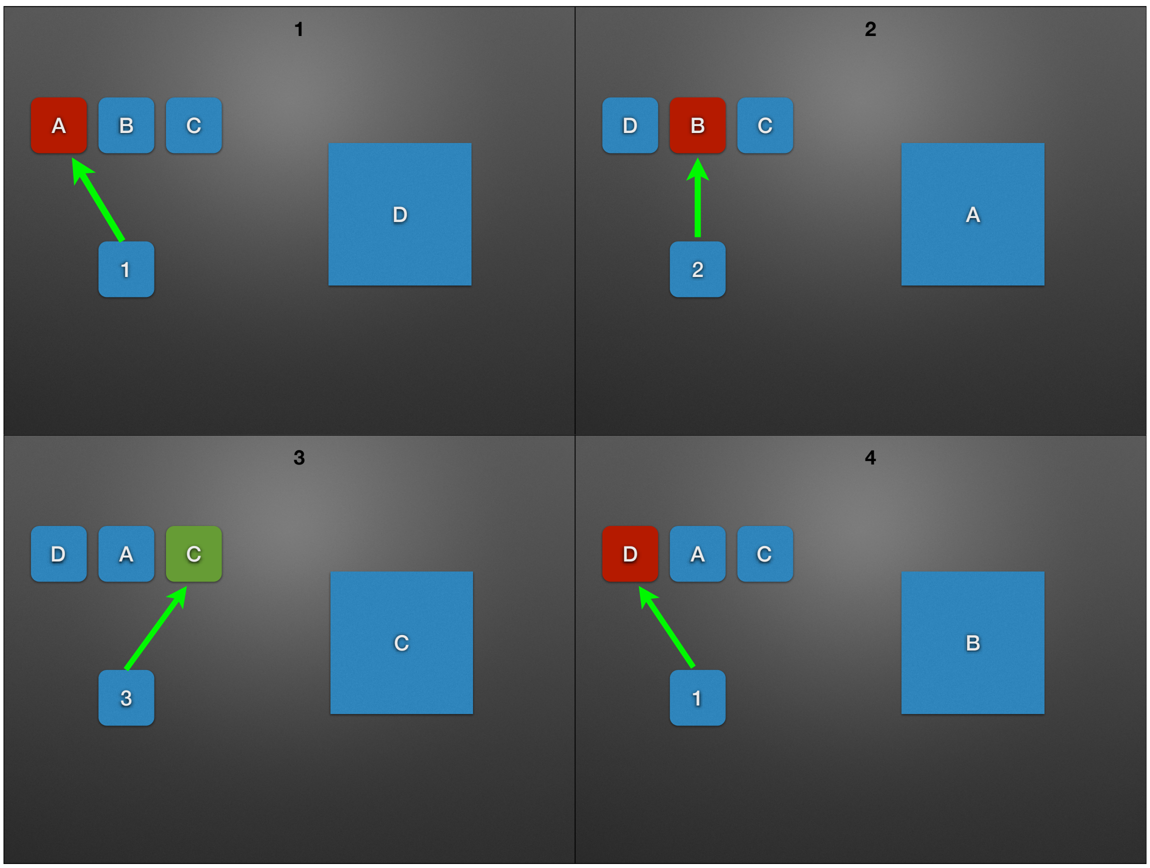
Strategies

- Your researcher will instruct you in the use of a specific strategy to use
- Please try to use this strategy even if you think there might be a better way to perform the task.
- Try to respond as quickly and accurately as possible, but do not worry if you make mistakes. The most important part of this experiment is to try to stick with the instructed strategy

These slides represent the rolling rehearsal strategy. On each trial the new item is compared to the left-most letter in the memory list. That letter is then removed from the list each item advances and the new letter is added to the right side of the list.



These slides represent the static rehearsal strategy. New letters replace the focused letter in memory. The pointer represents third previous letter to which the new letter is compared. This pointer moves sequentially on each trial.



E.2 Data and Statistics

Table E.1: ANOVA for sensitivity (d') for the full dataset of experiment 3

Error Subject	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Stimulus Distribution	1	5.20	5.20	0.53	0.4678	
Strategy	2	85.50	42.75	4.36	0.0146	0.06
Stimulus Distribution:Strategy	2	2.61	1.30	0.13	0.8756	
Residuals	136	1332.62	9.80			

Error: Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	11	11.36	1.03	1.99	0.0264	
linear	1	2.18	2.18	4.19	0.0408	
quadratic	1	3.70	3.70	7.10	0.0078	
cubic	1	0.54	0.54	1.04	0.3071	
Stimulus Distribution:Block	11	3.22	0.29	0.56	0.8596	
linear	1	0.10	0.10	0.19	0.6591	
quadratic	1	0.15	0.15	0.29	0.5933	
cubic	1	0.01	0.01	0.01	0.9129	
Strategy:Block	22	16.06	0.73	1.40	0.1012	
linear	2	2.23	1.12	2.15	0.1172	
quadratic	2	4.52	2.26	4.34	0.0132	
cubic	2	1.21	0.60	1.16	0.3134	
Stimulus Distribution:Strategy:Block	22	9.99	0.45	0.87	0.6330	
linear	2	0.77	0.39	0.74	0.4769	
quadratic	2	2.44	1.22	2.34	0.0963	
cubic	2	0.29	0.15	0.28	0.7562	
Residuals	1496	778.46	0.52			

Table E.2: ANOVA for d' values for the truncated dataset of experiment 3

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Stimulus Distribution	1	4.55	4.55	0.52	0.4720	
Strategy	2	56.10	28.05	3.21	0.0435	0.04
cond:Strategy	2	2.30	1.15	0.13	0.8768	
Residuals	136	1188.79	8.74			

Error: Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	8	7.62	0.95	1.96	0.0480	
linear	1	4.65	4.65	9.58	0.0020	
quadratic	1	0.24	0.24	0.49	0.4823	
cubic	1	0.09	0.09	0.19	0.6671	
Stimulus Distribution:Block	8	2.66	0.33	0.69	0.7051	
linear	1	0.04	0.04	0.08	0.7726	
quadratic	1	0.33	0.33	0.68	0.4091	
cubic	1	0.20	0.20	0.42	0.5181	
Strategy:Block	16	6.03	0.38	0.78	0.7142	
linear	2	0.30	0.15	0.31	0.7329	
quadratic	2	2.22	1.11	2.29	0.1017	
cubic	2	0.47	0.24	0.49	0.6148	
cond:Strategy:Block	16	7.60	0.48	0.98	0.4780	
linear	2	1.27	0.64	1.31	0.2708	
quadratic	2	2.23	1.12	2.30	0.1010	
cubic	2	1.30	0.65	1.34	0.2626	
Residuals	1088	528.27	0.49			

Table E.3: ANOVA for criterion values for the full dataset of experiment 3

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Stimulus Distribution	1	0.01	0.01	0.02	0.8929	
Strategy	2	1.81	0.90	1.88	0.1572	
Stimulus Distribution :Strategy	2	2.43	1.21	2.52	0.0842	0.04
Residuals	136	65.45	0.48			

Error: Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	11	2.17	0.20	2.43	0.0054	0.02
linear	1	1.21	1.21	14.82	0.0001	
quadratic	1	0.40	0.40	4.97	0.0259	
cubic	1	0.00	0.00	0.00	0.9722	
Stimulus Distribution :Block	11	0.96	0.09	1.08	0.3773	
linear	1	0.00	0.00	0.05	0.8182	
quadratic	1	0.00	0.00	0.00	0.9733	
cubic	1	0.00	0.00	0.04	0.8388	
Strategy:Block	22	1.68	0.08	0.94	0.5459	
linear	2	0.46	0.23	2.82	0.0599	
quadratic	2	0.17	0.08	1.03	0.3560	
cubic	2	0.01	0.01	0.08	0.9264	
Stimulus Distribution :Strategy:Block	22	2.45	0.11	1.37	0.1189	
linear	2	0.19	0.09	1.15	0.3174	
quadratic	2	0.15	0.08	0.95	0.3888	
cubic	2	0.34	0.17	2.11	0.1215	
Residuals	1496	121.75	0.08			

Table E.4: ANOVA for criterion values for the truncated dataset of experiment 3

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Stimulus Distribution	1	0.05	0.05	0.10	0.7565	
Strategy	2	2.03	1.01	2.11	0.1256	
Stimulus Distribution :Strategy	2	2.56	1.28	2.66	0.0736	0.04
Residuals	136	65.47	0.48			

Error: Subject:Block	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
Block	8	0.94	0.12	1.49	0.1547	
linear	1	0.20	0.20	2.51	0.1137	
quadratic	1	0.29	0.29	3.72	0.0542	
cubic	1	0.15	0.15	1.84	0.1749	
Stimulus Distribution :Block	8	0.44	0.06	0.70	0.6889	
linear	1	0.02	0.02	0.25	0.6143	
quadratic	1	0.05	0.05	0.68	0.4102	
cubic	1	0.00	0.00	0.02	0.8862	
Strategy:Block	16	0.85	0.05	0.68	0.8201	
linear	2	0.23	0.12	1.47	0.2312	
quadratic	2	0.05	0.03	0.32	0.7227	
cubic	2	0.00	0.00	0.01	0.9854	
Stimulus Distribution :Strategy:Block	16	1.68	0.11	1.34	0.1672	
linear	2	0.00	0.00	0.01	0.9922	
quadratic	2	0.27	0.13	1.70	0.1832	
cubic	2	0.09	0.05	0.59	0.5550	
Residuals	1088	85.65	0.08			

Table E.5: ANOVA - Match - Lure RT differences

Error - Subject	Df	Sum Sq	F value	Pr(>F)	η_p^2
Stimulus Distribution	1	282344.1	23.31	0.000	0.13
Strategy	2	198294.2	6.96	0.001	0.09
Stimulus Distribution:Strategy	2	74986.0	2.60	0.08	0.04
Residuals	136	1963618.9			

Error - Subject:Lure Position	Df	Sum Sq	F value	Pr(>F)	η_p^2
Lure Position	1	3275.4	0.47	0.50	
Stimulus Distribution:Lure Position	1	123685.0	23.13	0.000	0.14
Strategy:Lure Position	2	28855.1	2.55	0.08	0.04
Stimulus Distribution:Strategy:Lure Position	2	8586.7	0.77	0.46	
Residuals	136	755405.5			

Table E.6: ANCOVA - response times by lumpiness for n-1 lure trials

	Df	Sum Sq	F value	Pr(>F)	η_p^2
Stimulus Distribution	1	1648803.30	28.91	0.000	0.04
Strategy	2	394097.68	3.45	0.03	0.01
Lumpiness	1	36018.57	0.63	0.4271	
Stimulus Distribution:Strategy	2	1624931.95	14.24	0.000	0.04
Stimulus Distribution:Lumpiness	1	515038.19	9.03	0.003	0.01
Strategy:Lumpiness	2	13665.99	0.12	0.88	
Stimulus Distribution:Strategy:Lumpiness	2	68206.39	0.60	0.5503	
Residuals	658	37533214.87			

Table E.7: Experiment 3 - ANOVA tables of d' and c for control (non-trained) subjects

Sensitivity (d')

Error - Subject	Df	Sum Sq	F value	Pr(>F)	η_p^2
Stimulus Distribution	1	10.01	1.35	0.25	
Strategy	1	4.91	0.64	0.42	
Stimulus Distribution:Strategy	1	0.12	0.02	0.90	
Residuals	50	383.32.			

Error - Subject:Block	Df	Sum Sq	F value	Pr(>F)	η_p^2
Block	8	6.91	1.82	0.07	
Stimulus Distribution:Block	8	1.98	0.51	0.87	
Strategy:Block	8	2.42	0.55	0.82	
Stimulus Distribution:Strategy:Block	8	3.17	0.72	0.67	
Residuals	400	219.33			

Criterion (c)

Error - Subject	Df	Sum Sq	F value	Pr(>F)	η_p^2
Stimulus Distribution	1	0.17	0.25	0.62	
Strategy	1	0.44	1.04	0.32	
Stimulus Distribution:Strategy	1	0.30	0.73	0.39	
Residuals	50	20.58			

Error - Subject:Block	Df	Sum Sq	F value	Pr(>F)	η_p^2
Block	8	0.75	1.04	0.41	
Stimulus Distribution:Block	8	0.53	0.91	0.51	
Strategy:Block	8	0.48	0.68	0.71	
Stimulus Distribution:Strategy:Block	8	0.52	0.77	0.63	
Residuals	400	33.89			

Table E.8: Experiment 3 - ANOVA - Accuracy by lag position for control subjects

Error - Subject	Df	Sum Sq	F value	Pr(>F)	η_p^2
Stimulus Distribution	1	0.034	0.76	0.39	
Strategy	1	0.04	0.68	0.41	
Stimulus Distribution:Strategy	1	0.012	0.20	0.65	
Residuals	50	2.94			

Error - Subject:Lag Position	Df	Sum Sq	F value	Pr(>F)	η_p^2
Lag Position	9	0.98	23.14	0.00	0.36
Stimulus Distribution:Lag Position	9	0.03	0.72	0.68	
Strategy:Lag Position	9	0.10	2.67	0.005	0.06
Stimulus Distribution:Strategy:Lag Position	9	0.02	0.59	0.81	
Residuals	450	1.78			

Table E.9: Experiment 3 - ANOVA - Match-Lure response time differences for control subjects

Error - Subject	Df	Sum Sq	F value	Pr(>F)	η_p^2
Stimulus Distribution	1	7625.3	0.62	0.44	
Strategy	1	524.4	0.04	0.83	
Stimulus Distribution:Strategy	1	2556.2	0.20	0.63	
Residuals	50	628247.9			

Error - Subject:Lure Position	Df	Sum Sq	F value	Pr(>F)	η_p^2
Lure Position	1	42.67	0.01	0.93	
Stimulus Distribution:Lure Position	1	47226.81	8.67	0.004	0.14
Strategy:Lure Position	1	6710.70	1.20	0.28	
Stimulus Distribution:Strategy:Lure Position	1	3226.32	0.58	0.45	
Residuals	50	279672.10			