The Role of Uncertainty and Priority in Visual Working Memory

by

Aspen H. Yoo

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Psychology

New York University

September 2019

Wei Ji Ma

'What a useful thing a pocket-map is!' I remarked.

'That's another thing we've learned from your Nation,' said Mein Herr, 'map-making. But we've carried it much further than you. What do you consider the largest map that would be really useful?'

'About six inches to the mile.'

'Only six inches!' exclaimed Mein Herr. 'We very soon got to six yards to the mile. Then we tried a hundred yards to the mile. And then came the grandest idea of all! We actually made a map of the country, on the scale of a mile to the mile!'

'Have you used it much?' I enquired.

'It has never been spread out, yet,' said Mein Herr: 'the farmers objected: they said it would cover the whole country, and shut out the sunlight! So we now use the country itself, as its own map, and I assure you it does nearly as well.'

Lewis Carroll, from *Sylvie and Bruno Concluded*

Acknowledgements

i wish i could actually thank everyone who i feel scooted me along the trajectory that led to this point (i tried but it was inappropriately long). here are a few of the people who made my grad school experience what it was. 🥲

academically related ppl. my thesis readers, CRISTINA savin and DANIEL wolpert. you are both incredibly, intimidatingly smart people, but somehow manage to seem kind and approachable! thank you for the feedback on my thesis! my committee chair, MIKE landy, an ally even before getting accepted into nyu. im very thankful for you. LUIGI acerbi and EMIN orhan, who i've had the fortune of working with on previous iterations of the project that now is chapter 1. MARISSA evans, for collecting a fat chunk of that data!! ZUZANNA klyszejko, my coauthor for the projects in ch. 2 and 3. we were very good friends before and somehow tricked weiji and clay (or clayji as we coined) into letting us work together. also HELENA palmieri, for collecting most of the data in that project. GRACE hallenbeck, MASIH rahmati, and TOMMY sprague, all of whom are involved in chapter 4, thank you for letting me lay under your desk and for giving me pep talks, answering my dumb questions with a calm understanding, clarifying concepts that are confusing to me at an alarming depth and speed. RONALD van den berg and HONGSUP shin, my variable precision heroes, who i have been fortunate enough to work with in different ways. with ronald, i published a paper which is not in this thesis and, with hongsup, i worked with during a summer internship at arm (which was entirely unrelated to this line of work!!). BAS van opheusden, you definitely unstuck the recognition memory project with your magical physics juju. CLAY curtis, my "cool uncle" of an advisor, bc you aren't technically my advisor but have been supportive, thoughtful, responsive, and inspiring in ways that advisors should be. and of course, WEIJI ma, my actual advisor!

wow, i really hit the jackpot with you!! you are passionate, patient, understanding, helpful (most of the time), fun, thoughtful, and smart! you have been invested not only in my scientific growth but also in my personal growth (e.g., you are maybe the only person who has ever actually tried to get me to eat a tomato). and (maybe this is not a "good" trait, but it is exceptional...) you are probably the worst singer i've ever heard! in my life!

the massive amount of people who provide administrative, technical, and other support! especially SHENGLONG wang (hpc god), CLARKE (my best friend guard), SHEENA herrmann (og person helpful for graduation related stuff), DAVID park (current person helpful for graduation stuff), EVELYN rivera and JENNA rice (reimbursements), ROMAN zamishka (paychecks), ERIK cruz (subject payment reimbursements! efficient! nice nails!)

dear friends of mine who also happen to be in the ma lab: YANLI zhou, one of the most talented, well rounded people i've ever met. MAIJA honig, a large presence in the lab for many years. LUIGI acerbi, i will not forget our game of thrones nights or the one time you sang lady gaga poker face. WILL adler, maybe my first 'new york' friend, whatever the fuck that means. i have learned so many important life lessons from talking to and observing you and i love you so much for that. seBAStiaan caspar frans van opheusden, you used to annoy me when i first met you, but now you are one of my favorite people to talk to and be around. you are emotionally and academically and professionally supportive, you are kind and understanding in ways that sometimes go unappreciated, and you are silly! ANDRA mihali, wow my sister in the ma lab. you have been a damn rock for me many times. i love our snack trading, movie watching, meal eating, love discussing, astrology contemplating, weird-creature baking relationship. (other ma lab members i'd feel bad for not naming! YITIAN, DAVID halpern, EMIN orhan, ZAHY bnaya, HSIN-hung li, PEI PEI zhang, HEIKO schuett, XIANG li, IONATAN kuperwajs, JENN laura lee, ZHIWEI

li)

more friends! ZUZANNA klyszejko, my outdoor, pickle, fish queen. bonded by our love of pickled things and bathroom jokes (among other, more substantial things). GRACE hallenbeck, a presence throughout my graduate career, but recently has become key to my happiness at NYU. sometimes i came in only to say hey to you. EMILY boeke, my chillest most anxious friend. i loved our weekly lunches and more-than-weekly walks around, making weird paths around washington sq park because you liked to stray away from the building and i liked to stay close. the most considerate, supportive person. a wiz in the kitchen, on the piano, and in the canoe! oh how i pizza you! JENNIFER lenow, oh boy. i think the thing that really sparked what would become an extremely dynamic friendship was having to spoon for warmth in a literally freezing tent at some random ass thing zuzanna decided we were all going to i have probably learned as much about feelings and communication from you as i have learned about working memory in grad school!

undergraduate heroes. AMYBETH cohen, LAURA arce, iris blandon-gitlin aka IBG, ELISE fenn, and the rest of the MARC PROGRAM. i don't think i am certain of a lot of things, but i KNOWWWWW that without you i would not have gotten into grad school at all. i wish i could express my gratitude in more concrete ways than sending you an email here and there saying various unimaginative iterations of "hey thanks again for everything."

family (nonbiological and biological): ANNA kim, MINJU koo, ESTHER rhie, NICOLE an, ARAN kim, JON munoz, DAD, MASON, OPPA, MOM. i know none of you will read this, so all i want to say here is that i love you, you are everything to me, i can't imagine who i would be without the impact of your life on mine!

Preface

The work described in Chapter 1 is being prepared for journal submission. Complementary work that is not presented within this dissertation, but has served as great inspiration, was done in collabortion with Luigi Acerbi and Emin Orhan. On all of these projects, Weiji will be the sole senior author.

The work presented in Chapter 2 and 3 was jointly published with Zuzanna Klyszejko, Clayton Curtis, and Weiji Ma In *Scientific Reports* (Yoo, Klyszejko, Curtis, & Ma, 2018).

The work in Chapter 4 is still in progress, but will be written up for publication afterward. It is in collaboration with Grace Hallenbeck, Tommy Sprague, Masih Rahmati, Alfredo Bolaños, and Clayton Curtis, who will be the sole senior author.

Abstract

Visual working memory, the process involved in actively maintaining visual information over a short period of time, is essential for numerous everyday behaviors. A substantial amount of research has gone into characterizing the exact nature of one of the defining characteristics of visual working memory, its seemingly-counterintuitive limited capacity. Typically, these studies involve testing the memory for the value of a feature of some number of stimuli (e.g., the orientation of one of four oriented stimuli). However, more recent studies have shown that visual working memory holds more than just a point estimate of a stimulus feature, implying that the previous methods investigating visual working memory limits are underestimating the capacity and flexibility of this process. In my dissertation, I use psychophysical, computational, and neuroimaging methods to investigate how people maintain and use two additional pieces of information: an item's uncertainty, or the knowledge of the noise associated with its memory, and its priority, or behavioral relevance.

In the first chapter, I investigate uncertainty in working memory, showing that people maintain an accurate representation of item-specific uncertainty over a delay and use it optimally when deciding if a stimulus has changed in orientation. In the second chapter, I investigate how priority is used in working memory, specifically asking how we allocate our working memory resource across items with different priorities. I show that people allocate resources consistent with a loss-minimizing strategy.

In the third chapter, I investigate both priority and uncertainty, replicating and extending

findings from previous chapters. First, I show that people allocate resource in a way consistent with a loss-minimizing strategy, even when incentivized to use a different strategy. Second, I show that people maintain an item-specific representation of uncertainty, which isn't affected by priority information. Finally, I show that people use item-specific uncertainty optimally when placing a wager about the accuracy of their memory.

In the final chapter, I ask how priority is represented in the brain during working memory. I find that, in visual areas, priority is maintained through the amount of delay-period activity of the same neural populations maintaining each item's location.

Together, my studies demonstrate different ways in which working memory maintains and uses information that is helpful with interacting effectively with our environment.

Contents

List of Figures

List of Tables

0 Introduction

The best thing since sliced bread is a bunch of stuff between sliced bread

Bon Appétit

Working memory is responsible for actively storing and manipulating information for later use (Baddeley & Hitch, 1974; Baddeley, 2003). It is positioned between perception and cognition and is essential to behaviors as "simple" as integrating visual information across saccades (so we can maintain a steady mental representation of the world) and as "complex" as reading comprehension, problem solving, and decision making (Fukuda, Vogel, Mayr, & Awh, 2010; Conway, Kane, & Engle, 2003; Just & Carpenter, 1992). Thus, it is perhaps unsurprising that Schizophrenia, Parkinson's, and old age are associated with working memory deficits (J. Lee & Park, 2005; E.-Y. Lee et al., 2010; Park et al., 2002). The goal of this chapter is to describe and define working memory, specifically visual working memory, in relation to this dissertation.

0.1 Sensory, long-term, and working memory

To help define (visual) working memory as it is used in this thesis, I want to first discuss how it is different from sensory memory (also known as iconic memory), which it follows, and long-term memory, which it precedes.

0.1.1 Sensory memory

A seminal study published by Sperling in 1960 demonstrated the difference between the fragile but high-resolution sensory memory available immediately after perception, and the more robust but lower-capacity working memory. Participants were briefly presented a display containing 3-12 alphanumeric characters, arranged in one to three rows (e.g., one row of 3 or 6, three rows of four). There were two response conditions, whole report and partial report. In the whole report condition, participants were instructed to report

as many characters as they could from the entire display. In the partial report condition, participants were probed to report as many characters from a subset of the entire display. For example, a high, medium, or low auditory tone would be played and the participant would respectively recall the remembered characters from the top, medium, or bottom row of the display. The tone, and thus the row to report, was randomly selected. The amount of time between the display presentation and the probe (i.e., the delay) varied as well. Sperling found that, in the whole report condition, participants were able to recall at most 4-5 items, regardless of the total number of characters in the display. In the partial report condition, participants were able to report 3-4 items when the response tone was played directly after stimulus presentation. Because participants were unaware of which row would be probed for response (and thus would perform equally well for all rows), this result suggested that participants had 9-12 items available in memory immediately after viewing the display. However, this sensory memory was fragile and very short lived, vulnerable to interference of a post-stimulus mask and declined quickly in the first second after stimulus presentation. After about a second, performance in both conditions was equivalent and relatively stable across delay times, suggesting a shift into a more stable, but lower-capacity memory system.

0.1.2 Long-term memory

While working memory is severely capacity limited, long-term memory for visual stimuli seems to have no bounds. In a 1973 study published by Standing, participants were tested on their long-term memory for images and words. Participants studied between 20 and 10,000 images, knowing they would be tested on their memory for a subset. Approximately two days after the study session, participant were tested on their memory for a

subset of images with a recognition test. Standing found that participants' memory for images increased logarithmically with the number of images studied, implying that there is no upper bound to the number of visual stimuli we can remember. Results of similar studies suggest that this high performance reflects a relatively rich memory for perceptual details of the stimulus, not simply category relations (Brady, Konkle, Alvarez, & Oliva, 2008; Konkle, Brady, Alvarez, & Oliva, 2010). Brady and others (2013) additionally found that the fidelity of visual information in long-term memory is comparable to that of visual working memory, suggesting that information acquires no additional loss when being consolidated to long-term memory.

0.1.3 Working memory

It is clear, when contrasting different memory processes, that working memory is a process that actively maintains a limited amount of information over a short period of time. In my dissertation, I assume the "limited amount of information" is realized through a continuous, stochastic resource, which I describe further in Section 0.2.3 below.

What are the neural mechanisms that support the encoding and maintenance of WM items? A prominent theory was that memory is maintained through persistent activity in the lateral prefrontal cortex during the delay (e.g. Fuster & Alexander, 1971; Funahashi, Bruce, & Goldman-Rakic, 1989). Seminal work by Funahashi and others (1989) found, in a monkey oculomotor delayed-response task, that a significant proportion of neurons recorded in monkey dorsolateral prefrontal cortex sustained neural activity above baseline during the memory delay. Furthermore, they found neurons that were sensitive to space, consistently having greater activity for one target location than another.

More recently, delay-period activity associated with working memory has been shown to be distributed throughout the brain, depending on the task demands and the stimulus being maintained in working memory, supporting a more "sensory recruitment" model of working memory (e.g. Curtis & D'Esposito, 2003; Postle, 2006; D'Esposito & Postle, 2015; D'Esposito, 2007; Y. Xu, 2017; Harrison & Tong, 2009). The premise of this model is that the brain maintains information with the same areas that encoded that information, rather than copying that information. Crucially, this means that working memory for visual items takes advantage of the system already designed to hold that information at high resolution. This revision of the original model of working memory maintenance in the brain posits that a high-fidelity representation of sensory areas is maintained in visual areas, which is modulated by goal-directed information represented in prefrontal cortex (Curtis & D'Esposito, 2003; Sreenivasan, Curtis, & D'Esposito, 2014).

0.2 VWM encoding: data and models

In this section, I will discuss data and behavioral models that provide insight into how working memory representations are encoded. First, I will discuss the "slots" model, which posits that memory is object-based, and limited by how many objects can be stored within it (e.g. Luck & Vogel, 1997; Cowan, 2001). I will then discuss data that the original slots model is not able to account for, then introduce the "resource" models, which posit that working memory has a limited amount of information that is shared across memoranda (e.g. Wilken & Ma, 2004). Finally, I will discuss how the behavioral relevance of different items affects how they are encoded.

0.2.1 Slots models

Different instantiations of slots models have been around for a while, like Miller's or Cowan's respective "magical" 7 ± 2 or 4 (Miller, 1956; Cowan, 2001). The idea behind this theory is that there are a limited number of slots, which are responsible for maintaining memories. If an item makes it into a slot, it is remembered perfectly. If it doesn't make it into the slot, it is not remembered at all. A metaphor sometimes used to describe this model is having three to four juice boxes to distribute to people. People either receive an entire juice box or no juice at all. This model could explain the clear decline in performance when increasing set sizes beyond 3 or 4.

In 1997, Luck and Vogel published a seminal study using a change detection paradigm. In change detection tasks, participants briefly view a set of items, remember them over a delay, view another set of items, and decide whether the second set of items is the same as the first in the relevant feature(s). Typically, if there is a change in the second set, then the change is in one item. In this task, participants completed change detection tasks with colored squared, oriented lines, or lines with both orientation and color. Importantly, they varied the number of stimuli to be remembered (i.e., the set size) from 1 to 12. This study found that performance was equally good for set sizes 1-3, but declined from 4-12. Additionally, they found that the same performance was achieved whether the stimuli had only one relevant feature (e.g., color or orientation) or multiple relevant features to remember (e.g., color and orientation). Luck and Vogel concluded from these results that our memory was composed of several object-based representations, similar to "chunks" in verbal working memory. This means that people can hold several items in mind, irrespective of the complexity of the item.

0.2.2 Data incompatible with the slots framework

While the slots model was able to capture some set size effects of data, there were trends in the data it could not account for. Specifically, there were a number of studies demonstrating a tradeoff between the number of items one could remember with the precision with which they remembered them. For some of these tasks, the set size tested was well within the capacity limits laid out by the slots model, indicating that memory was not actually object-based (and thus constrained only to one's ability to chunk things into "objects"), but information-based (only able to hold a certain amount of information across all features of all items). Here, I summarize some of this data.

Phillips (1974) demonstrated in a change detection task that the complexity of a single item affected its change detection performance. Alvarez and Cavanaugh (2004) demonstrated that capacity estimates for different item types (line drawings, cubes, random polygons, Chinese characters, letters, colors) were directly related to its complexity, finding that the estimated capacity was lower for more complex items. These results suggest that, if working memory is made up of a few slots, then these slots are not truly-object based but themselves have some capacity limit.

In 2004, Wilken and Ma demonstrated that there was a tradeoff between number of items remembered and the precision with which they are remembered. In a delayed estimation task, participants were instructed to remember the feature value of N stimuli (i.e., the color of 2, 4, 6, or 8 colored squares, the orientation of 2, 3, 4, or 5 oriented Gabors). After a 1500ms delay, one of the N stimuli was probed, and participants reported the value of that probed stimulus as accurately as possible. Unlike previous recall tasks, which used discrete stimuli like alphanumeric characters (e.g. Sperling, 1960), these stimulus features

were continuous. Thus, the memory estimation errors could be aggregated across trials to make an error distribution, where the center was the true stimulus value and the width corresponded to some measure of a person's memory precision. These results found that, for higher set sizes, the distributions were wider, illustrating a tradeoff between the number of items remembered and the precision with which they are remembered.

Fougnie and others (2010) showed that items of different complexity could be estimated to have the same capacity, but the precision with which the items are remembered would differ. Specifically, they found that items with both color and orientation information had the same capacity estimates as items with only color or orientation information, in agreement with the results of Luck and Vogel (1997). However, an important additional result that previous authors didn't show was that memory for items with multiple features was less precise than items with just one feature.

0.2.3 Resource models

Because of the inability of the slots model to account for this data, a new family of models was suggested, which are broadly referred to as "resource" models (e.g. Wilken & Ma, 2004; Fougnie, Suchow, & Alvarez, 2012; van den Berg, Shin, Chou, George, & Ma, 2012; Sims, Jacobs, & Knill, 2012; Bays & Husain, 2008; Oberauer & Lin, 2017). The fundamental idea behind these models is that there is a continuous, limited resource that all items in memory must share. This is akin to having pitcher of juice that one must distribute amongst all people. This juice could be allocated flexibly: all juice could be distributed to one item, juice could be distributed equally across items, juice could be distributed unequally across a subset of items. The amount of juice was not necessarily fixed, some versions allowing the total amount of resource to vary with set size (Bays & Husain,

2008; van den Berg et al., 2012; Keshvari, van den Berg, & Ma, 2013; van den Berg & Ma, 2018; Mazyar, van den Berg, & Ma, 2012; Elmore et al., 2011) or on a trial-to-trial basis (Fougnie et al., 2012; van den Berg et al., 2012).

In my thesis, I primarily build models consistent with the Variable Precision (VP) model described by van den Berg, Shin, and others (2012). This model assumes that there is a total amount of resource available to an observer, defined as total memory precision, which varies on a trial-to-trial basis. This assumption, as well as its lack of assumptions about resource allocation across items, were two key reasons why we chose to use this model. In Chapter 1, I quantitatively test whether variable precision is a reasonable assumption to make in human data, and find strong support for this model, corroborating previous results (van den Berg, Awh, & Ma, 2014). For Chapter 2 and 3, I continue to assume that precision is variable. The assumption of variable precision is important for Chapter 3, and the lack of constraints about resource allocation is critical to Chapter 2 and 3.

In this model, the total memory precision is operationalized as Fisher information. We model an observer's noisy memory of an item as a sample from a Gaussian or von Mises (for stimuli that live in a circular space) distribution. To account for the observed stochasticity in memory precision across items and trials, the precision of this Gaussian or von Mises is itself a random variable. In this VP model, we assume that the precision is drawn from a Gamma distribution.

0.2.4 The role of priority in VWM

One behavioral effect that I will be paying particular attention to in this dissertation is that people can remember certain items better, often at the cost of the memory for other

items. Experimenters elicited this effect by manipulating the behavioral relevance of the items either through the probability of it being probed (e.g. Bays, 2014; Dube, Emrich, & Al-Aidroos, 2017; Emrich, Lockhart, & Al-Aidroos, 2017; Klyszejko, Rahmati, & Curtis, 2014; Zhang & Luck, 2008; Gorgoraptis, Catalao, Bays, & Husain, 2011) or by monetary reward (Klyszejko et al., 2014). For example, in the 2014 Bays paper, a location was cued before stimulus presentation, indicating that the item that would appear in that location was three times more likely to be probed for response than the remaining item(s). In this dissertation, I define priority as behavioral relevance. While this bears some similarity to the "priority" of priority maps theory (e.g. Thompson & Bichot, 2005; Serences & Yantis, 2006; Bisley & Goldberg, 2010), which posits that the priority of an item is a function of both the bottom-up salience and top-down importance, I restrict the meaning in this dissertation to be purely top-down. In Chapter 2 and 3, I investigate what strategies people use when allocating resource across items with different priorities.

How might priority be represented in the brain? Perhaps it is maintained through neural population gain. Human fMRI studies have shown that attention increases the neural gain of populations encoding the attended item's location (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; Buracas & Boynton, 2007; Gandhi, Heeger, & Boynton, 1999; Gouws et al., 2014; Jerde, Merriam, Riggall, Hedges, & Curtis, 2012; Serences & Yantis, 2007; Somers, Dale, Seiffert, & Tootell, 1999; Rahmati, Saber, & Curtis, 2018; Sprague, Itthipuripat, Vo, & Serences, 2018; Saber, Pestilli, & Curtis, 2015; Nobre et al., 2004). Additionally, population coding models have successfully accounted for prioritization effects in error through modulating neural gain (Bays, 2014). In Chapter 4, I investigate whether an item's priority is represented through the amplitude of the same populations encoding its location.

0.3 VWM decisions: data and models

0.3.1 Bayesian decision theory

In the previous section, I described two families of models of how working memory is encoded. However, those models don't describe how the memories are then used to make a decision. For that, I use Bayesian decision models, which provide a flexible, interpretable, and generalizable framework to study the working memory decision process. Bayesian decision models are particularly useful in cases where the observer is trying to make a decision without full knowledge of task-relevant information. In working memory, this is often because people don't remember information perfectly. Bayesian decision models have two components: Bayes rule, which formalizes how observers compute probabilities about states of the world, and a cost function (also referred to as gain, loss, or objective function), which describes how observers should use the probabilities in a decision. While Bayesian decision theory describes how an observer should behave in order to maximize performance, different components of the model can be easily substituted with incorrect beliefs or suboptimal use of information, and thus provides a good template for building models with "imperfectly optimal observers" (Maloney & Zhang, 2010) or "model mismatch" (Orhan & Jacobs, 2014; Beck, Ma, Pitkow, Latham, & Pouget, 2012; Acerbi, Ma, & Vijayakumar, 2014).

The first component of Bayesian decision models, Bayes rule, allows the observer to calculate beliefs over different states of the world from their observations (or in this dissertation, their memories) and prior knowledge. More concretely, a Bayesian observer can compute posterior probabilities of states of the world as a function of the prior (the probabil-

ities of the states of the world) and the likelihood (the probability of having their current memory given different states of the world). The observer combines this posterior with the cost function to calculate the expected cost of each action. With this information, the optimal observer chooses the action that produces the lowest expected cost.

In my dissertation, I investigate how (if at all) people are using uncertainty in their working memory decisions. Intuitively, uncertainty is how trustworthy we believe our memory is; technically, uncertainty is the believed width of the likelihood or posterior distribution over the stimulus value.

Uncertainty and noise are not synonymous. Noise refers to the imprecision at the encoding process, which could be due to a noisy stimulus or poor memory. Uncertainty refers to the ambiguity associated with the stimulus value during the decision stage. Noise and uncertainty may be completely unrelated. For example, consider looking at a bunch of dead leaves on the ground. There may be uncertainty about which leaf segments correspond to the same leaf because of occlusion, despite there being no sensory noise.

However, often times noise contributes to uncertainty. For example, viewing a low-contrast Gabor will results in higher sensory noise than a high-contrast Gabor. For the optimal Bayesian observer, this noise will affect the uncertainty over the orientation of the Gabors; the uncertainty associated with the orientation of the low-contrast Gabor will be higher than that of the high-contrast Gabor. However, it is possible for the observer's uncertainty to not match that of the Bayes-optimal observer, which we will colloquially refer to as "inaccurate." In my dissertation, I manipulate noise through sensory reliability (Ch. 1) or priority (Ch. 2, 3).

0.3.2 The role of uncertainty in VWM

In perception, there have been numerous studies demonstrating optimal use of uncertainty in a variety of tasks (e.g. van Beers, Baraduc, & Wolpert, 2002; Trommershäuser, Maloney, & Landy, 2003; Ernst & Banks, 2002; Alais & Burr, 2004; Kording, 2007; Knill & Pouget, 2004). However, its role in working memory tasks has not garnered the same amount of attention. Here are three reasons why uncertainty in working memory was not considered until recently. First, categorization tasks, such as change detection, are trivial from a slots model framework. If working memory was an all-or-none process, making a change detection decision would be as simple as saying the stimulus has not changed only if it is exactly what you remember and saying it changed otherwise. In other words, there is no uncertainty in the slots model framework. In a resource framework of working memory, change detection becomes a signal detection or Bayesian inference problem (Wilken & Ma, 2004), in which uncertainty is fundamental to maximizing performance. This is because the observer must now to weigh their memory by how noisy it is. Second, the magnitude of changes in traditional change detection tasks was so large, for example from blue to yellow, that use of uncertainty would not necessarily improve performance. Uncertainty assists an observer when making a categorical decision like change detection only if the measured amount of change could be reasonably attributed to either the actual change or memory noise. Third, some tasks commonly used to study working memory do not actually require use of uncertainty to maximize performance. Examples of these tasks are change discrimination and delayed estimation tasks, both of which only require use of a point estimate of the value of a memory to make a performance-maximizing decision.

Recently, there has been some research investigating the role of uncertainty in working

memory. There is some evidence that people maintain a sensible representation of uncertainty. People's memory reports are more accurate in trials with higher confidence reports (Rademaker, Tredway, & Tong, 2012). Even within a trial, people are able to choose a better remembered item to report, suggesting an item-specific representation of their memory noise (Suchow, Fougnie, & Alvarez, 2017). However, other studies have indicated that accuracy and confidence need not be related, suggesting our uncertainty may not be accurate (Bona & Silvanto, 2014; Adam & Vogel, 2017). On the other hand, there is research indicating that observers can use uncertainty cues in a change detection task, but did not investigate whether participants maintained that uncertainty information in VWM (Keshvari, Berg, & Ma, 2012). The goal of Chapter 1 is to answer whether people maintain and use an accurate representation of uncertainty in a change detection task.

0.4 Dissertation outline

In my thesis, I discuss specifically how priority and uncertainty are used to facilitate working memory representations and decisions. In Chapter 1, I investigate the role of uncertainty in working memory. Specifically, I ask whether uncertainty information is maintained and used in an orientation change detection task. I find, through computational modeling, that people maintain an accurate representation of their uncertainty and use it probabilistically when making a change detection decision.

In Chapter 2, I investigate how priority is used to facilitate the encoding of working memory representations. Specifically, I ask if people flexibly allocate resource across multiple levels of priority and, if so, how. I use a four-item delayed estimation task where each item has a different priority. I find, through computational modeling, that people allocate re-

source in order to minimize their estimation error across the experiment.

In Chapter 3, I investigate uncertainty and priority together, asking if we can replicate and generalize the results of previous chapters in new contexts. Specifically, I ask if priority affects uncertainty representations, if people use uncertainty optimally when making working memory decisions, and if people would change their encoding strategy based on task demands. I find remarkably consistent results: people have an accurate item-specific representation of uncertainty, people use this uncertainty optimally when making decisions, and people allocate resource in order to minimize error (although this strategy is myopic in this experiment).

Finally, in Chapter 4, I investigate the representation of priority in the brain during a working memory delay. Specifically, I use fMRI to ask if the priority of an item, in a spatial delayed estimation task, is represented in the same populations maintaining its location. I find evidence of this representation in visual areas, but not in frontal areas. These results provide direct evidence of gain as the neural mechanism of priority, and thus perhaps precision, in sensory areas. Additionally, it demonstrates the different representational structures, and thus perhaps functional roles, of different brain areas.

1 The role of uncertainty in change detection

Perhaps the earth is floating,
I do not know.
Perhaps the stars are little paper cutups
made by some giant scissors,
I do not know.
Perhaps the moon is a frozen tear,
I do not know.
Perhaps God is only a deep voice,
heard by the deaf,
I do not know.

Anne Sexton, *The Poet of Ignorance*

1.1 Introduction

Traditional slot theories of visual working memory (VWM) have described remembering something as an all-or-nothing process (e.g., Luck & Vogel, 1997). If an item makes it into a slot, it is remembered perfectly. If it does not make it into a slot, it isn't remembered at all. According to this model, categorization tasks such as change detection are trivial. Participants would be absolutely certain when deciding about a remembered items and guessing otherwise.

More recent theories have described memories as noise-corrupted representations of memoranda. In this framework, VWM is a limited resource that can be flexibly allocated to any number of stimuli or stimulus features (e.g., Bays & Husain, 2008; Fougnie, Cormiea, Kanabar, & Alvarez, 2016). The models in this resource framework explain human data better than previous theories (van den Berg et al., 2012; Fougnie et al., 2012; Bays & Husain, 2008; Ma, Husain, & Bays, 2014). In a resource model framework, change detection is no longer trivial. Because each memory is an imperfect measurement of the actual stimulus, change detection becomes a signal detection problem. In order to perform optimally, an observer must maintain the uncertainty information associated with each memory (Wilken & Ma, 2004). While optimal use of uncertainty has been established in the perceptual literature (e.g. van Beers et al., 2002; Trommershäuser et al., 2003; Ernst & Banks, 2002; Alais & Burr, 2004; Kording, 2007; Knill & Pouget, 2004), its role in working memory tasks is uncertain (pun intended). Do people maintain and use uncertainty? Do people perform optimally in VWM change detection tasks?

Keshvari and others (2012) demonstrated that people use uncertainty cues optimally in a change detection task. In their experiment, participants remembered the orientations

of four ellipses over a one second delay. After the delay, the ellipses were presented again briefly. On half of the trials, the ellipses were identical to the first presentation. On the other half of the trials, the orientation of one of the ellipses changed. Participants indicated with a button press whether they believed there was a change or not.

There were two important experimental departures from traditional change detection tasks, both of which allowed the experimenters to test how accurately people represent their uncertainty. First, the change in orientation on every trial was randomly sampled from a uniform distribution, and thus could be of any magnitude. Previous tasks typically used a fixed, large change. Like in previous tasks, detecting a large change is relatively easy; it is highly unlikely that you would measure such a change from memory noise alone. However, small measured changes could be a result of an actual physical change or memory noise. Uncertainty is essential for disambiguating these two potential states of the world.

Second, each stimulus could independently have "high" or "low" sensory reliability. In this study, we manipulated it through ellipse elongation; a longer, skinnier ellipse provided higher reliability orientation information than its shorter, wider counterpart. If participants' uncertainty was accurate (which is to say, if their uncertainty reflected the uncertainty of the optimal Bayesian observer), then their uncertainty on average would be lower for high reliability items. In order to perform optimally on this change detection task, participants would have to weigh their memories of each of the ellipses by its item-specific uncertainty. Thus, this manipulation allowed experimenters to distinguish between participants who do and do not have an item-specific representation.

Keshvari and others found that a model with the following three characteristics fit the data best: doubly-stochastic memory, as defined through the Variable Precision model

(van den Berg et al., 2012); observers who had an accurate, item-specific uncertainty; and observers who optimally combined information to make their decision. This study showed strong evidence for the use of uncertainty in working memory tasks, demonstrating that change detection was a form of probabilistic inference.

However, the notion of memory is largely absent from this paper. Memory is only mentioned briefly in the discussion when explaining the complexity of the change detection decision process; the authors recognize that the orientations must be maintained over the delay. However, they don't discuss whether *uncertainty* is maintained over the delay. In fact, it is possible that uncertainty was not maintained in working memory. The main manipulation of uncertainty in this task was through the noise induced by having different ellipse reliabilities. Because ellipses were presented with the same reliabilities on the second delay, participants could have used that information as a heuristic for uncertainty instead of maintaining any representation.

The results about whether people can maintain uncertainty is a bit mixed. Most of these studies probe confidence, an explicit report related but not identical to uncertainty. Some of these studies find evidence of a sensible representation of uncertainty. For example, Rademaker and others (2012) demonstrated that people's memories were more precise on trials that they were more confident in. Suchow and others (2017) demontrated, in two different tasks, that people can reliably choose the better remembered item out of multiple memoranda, suggesting an item-specific representation of working memory precision.

Others show evidence of the opposite (e.g. Bona & Silvanto, 2014; Adam & Vogel, 2017). Bona and Silvanto found that making two orientation change discrimination judgments in a row significantly decreased confidence, but did not necessarily decrease accuracy. Adam and Vogel demonstrated a disconnect between participants' feeling of remembering and

actually remembering. In several experiments, participants would sometimes fail to remember anything despite reporting being "on task." Thus, it is still unclear from these confidence studies whether people maintain a representation of uncertainty.

In this chapter, we ask whether people maintain and use an accurate representation of uncertainty in an orientation change detection task. We chose to investigate this question by subtly changing the experimental paradigm in Keshvari et al., 2012. In the first experiment, we replicate the experiment and models finding that people can use uncertainty information optimally when that information is available to them at decision. We then modify the experimental paradigm such that uncertainty information must be maintained in order to be used and ask if people are able to maintain and use memory uncertainty. We find remarkably consistent results, suggesting that people maintain an accurate representation of uncertainty over the delay, and use it when making change detection decisions.

1.2 Experiment

1.2.1 Experimental methods

1.2.1.1 Stimuli

Stimuli were four, light-grey, oriented ellipses on a medium-grey background. Each ellipse could be "long" or "short," to provide respectively higher or lower reliability information regarding the orientation of the ellipses. All ellipses had an area of 1.19 degrees of visual angle (dva). The long ellipse had an ellipse eccentricity of 0.9, such that the major axis and minor axis was 1.02 and 0.37 dva, respectively. The ellipse eccentricity of the short

ellipse was determined separately for each participant to equate performance (details in Procedure).

On every trial, a stimulus display consisted of four ellipses. The probability of each ellipse being a long ellipse was 0.5, independent of the reliability of the other ellipses. The location of the first ellipse was drawn from a uniform distribution between polar angles 0° and 90°. Each ellipse after that was placed such that all ellipses were 90° apart on an imaginary annulus that was 7 dva away from fixation. Afterward, the x- and y- location of the ellipses were independently jittered -0.3 to 0.3 dva. In one experiment, there were additionally oriented line stimuli, which were set to have approximately the same area as the ellipses.

1.2.1.2 Participants

Thirteen participants (11 female; $M = 21.1, SD = 2.5$) completed both experiments. All participants had normal or corrected-to-normal vision. Participants were naive to the study hypotheses and were paid $12/hour and a $24 completion bonus. We obtained informed, written consent from all participants. The study was in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of New York University.

1.2.1.3 Procedure

Participants completed two experiments. In one experiment, a trial began with a fixation cross presented for 1000 ms. Four ellipses were presented for 100 ms, followed by a 1000 ms delay, then by the second stimulus set presentation for 100 ms. On half of the trials,

all ellipses in the second stimulus presentation were identical to the ellipses in the first stimulus presentation. On the other half of the trials, one ellipse changed in orientation; this change was drawn from a uniform distribution, so change of any magnitude had equal probability. Each ellipse had equal probability of containing the change. The participant indicated with a button press whether they believed there was an orientation change between the two displays.

In the second experiment, the stimuli in the second stimulus set presentation were oriented lines rather than ellipses. The task was otherwise identical. We refer to the experiments with ellipses and lines in the second presentation as the "Ellipse" and "Line" experiments, respectively. Examples of a trial in the Ellipse and Line experiments is illustrated in Figure 1.1.

Participants completed both experiments over six one-hour sessions. They began their first session with a Practice block, used to ease the participants into the task. They then completed 2000 trials of each experiment, preceded by a Threshold block to set the "short" ellipse reliability for each experiment. Participants completed all of one experiment before completing the other, and the order was counterbalanced across participants.

The Practice block consisted of 256 trials. The stimulus presentation time decreased throughout the course of the Practice block, from 333 ms to 100 ms, in 33 ms increments every 32 trials. The stimuli in the second stimulus presentation corresponded to the experiment that the participant completed first. For example, the stimuli in the second presentation were lines if the participant completed the Line experiment first.

The Threshold block consisted of 400 trials, and was used to set a short ellipse reliability level for each experiment. Unlike the actual task, the reliabilities of all ellipses on each

trial were the same, but changed on a trial-to-trial basis. The second stimulus presentation set were either ellipses (of the same reliability) or lines, corresponding to the experiment following it. A cumulative normal psychometric function was fit to the accuracy as a function of ellipse reliability, and the "short" ellipse reliability was set as the value that corresponded to a predicted 65% accuracy. If the ceiling performance of the participant was estimated to be less than 75%, the Threshold block was repeated. If the psychometric function could not estimate an ellipse reliability for which performance would hit 65% after the second try, the participant was excluded from the experiment.

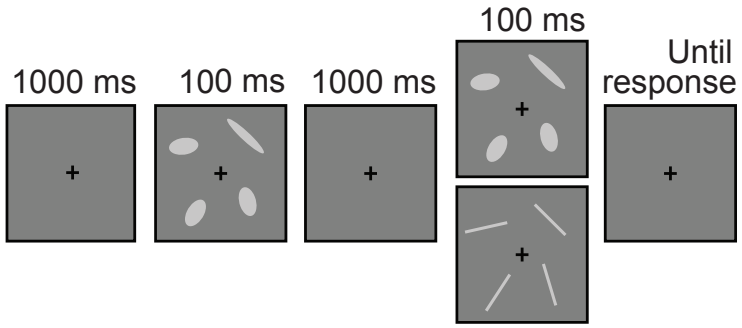

Figure 1.1 Trial sequence. Participants viewed a fixation cross, four ellipses were presented (here showing 1 high-reliability ellipse and 3 low-reliabilty ellipses), maintained them over a delay, saw four stimuli again, and reported whether they believed there was an orientation change or not. In the Ellipse experiment, ellipses of the same reliability as the first presentation were presented during second display. In the Line experiments, lines were presented instead of ellipses, to avoid providing cues to the precision with which the first items were maintained.

1.2.2 Experimental results

Data for both experiments are visualized in Figure 1.2. Figure 1.2A show the proportion report "change" as a function of magnitude of change (x-axis) and number of high-reliability items (N_{high}, shown through different colored lines), for the Ellipse experiment on the left and the Line experiment on the right. We conducted a two-way repeated-

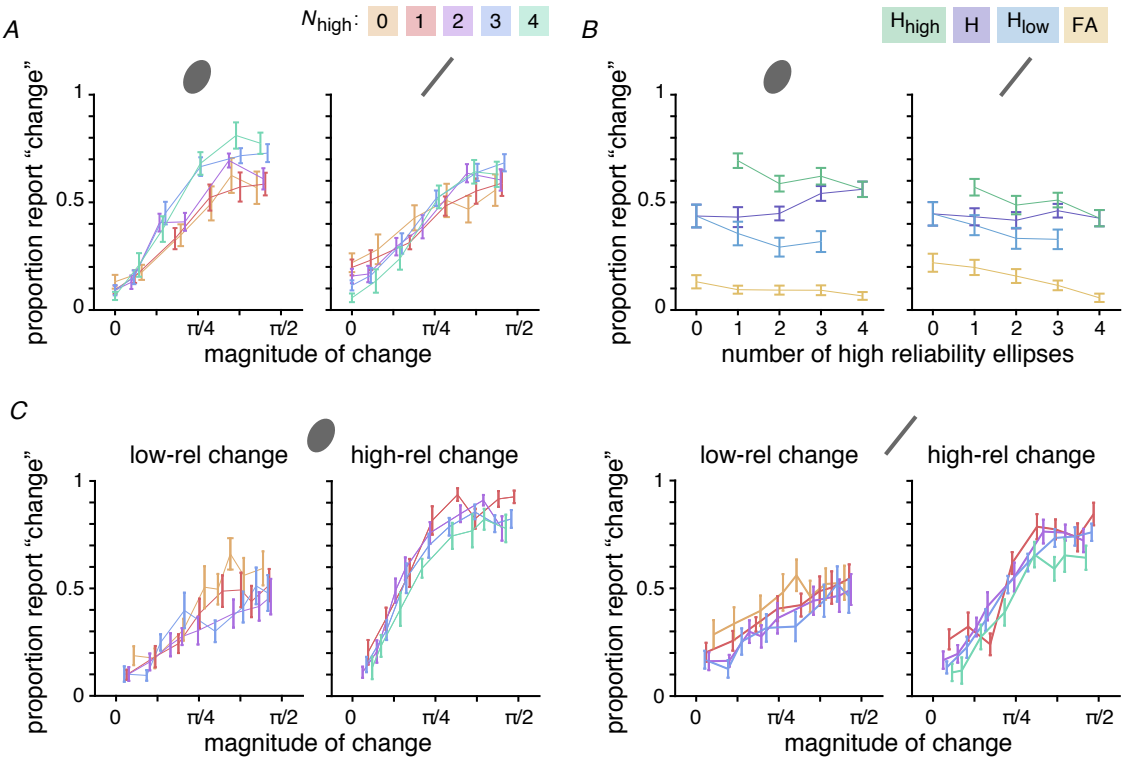

Figure 1.2 Behavioral data. *A*. proportion report "change" as a function of magnitude change for Ellipse (left) and Line (right) experiments, conditioned on how many high-reliability stimuli were presented (N_{high}, illustrated through different colors). *B*. proportion report "change" as a function of number of high-reliability ellipses for Ellipse (left) and Line (right) experiments, conditioned on whether there was no actual change (false alarm, FA, gold), a change in a low-reliability ellipse (H_{low}, blue), a change in a high-reliability ellipse (H_{high}), or a change in any ellipse (hit, H, purple). Note that the hits, not conditioned on reliability of stimuli, are a weighted function of the conditioned hits functions. *C*. proportion report "change" as a function of magnitude of change, broken up by whether the change occurred in low- or high-reliability ellipses. The left two plots correspond to psychometric functions with low- and high-reliability change in the Ellipse experiment. The right two plots correspond to the psychometric functions with low- or high-reliability change in the Line experiment.

measured ANOVA, with experiment and N_{high} as factors. The effect of N_{high} was approaching significance ($F(1.38, 16.51) = 3.37, p = 0.07$), after making a Greenhouse-Geisser correction to the violation of sphericity. There was a nonsignificant effect of experiment, but a significant interaction between N_{high} and experiment ($F(2.12, 25.39) = 6.32, p = 0.006$). These reflect the qualitatively observable effect of N_{high} on proportion respond "change" for the Ellipse experiment only; people report "change" more when there are more reliable ellipses. This seeming lack of effect of N_{high} in the Line experiment may seem like uncertainty is not being used in the task. However, this qualitative difference can be caused by things besides a lack of use of uncertainty. For example, it could be caused by the change in stimulus type from the first to second display. It could also be caused by the decrease in uncertainty in every item from having an extremely high-reliability orientation stimulus in the second display. A process model can distinguish whether this qualitative difference is due to a difference in strategy.

Figure 1.2B shows the false alarm rates and hit rates, broken up and averaged across high- and low-reliability items, for different N_{high}, for the Ellipse experiment on the left and the Line experiment on the right. Again, we test effects of number of high-reliability items N_{high} and experiment on these variables with a two-way rmANOVA. The false alarm rate for the Line experiment ($M = 0.14, SEM = 0.03$) was significantly higher than that of the Ellipse experiment ($M = 0.09, SEM = 0.02, F(1, 12) = 6.4953, p = 0.02$). The greater false alarm rate in the Line condition could be due to people mistaking the change in stimulus as a change in orientation. The significant effect of N_{high} ($F(4, 48) = 18.21, p < 0.001$) and a significant interaction between the two ($F(1.94, 23.36) = 4.94, p = 0.04$) corroborates this intuition. Because the high-reliability items provide more orientation, they provide more evidence of "no change," thus making more correct responses as the number of high-reliability ellipses increase. This may result be even larger for the Line experiment because

of the stimulus change; the change from a high-reliability ellipse to a line may not feel as jarring as a difference as the change from a low-reliability ellipse to a line.

There was a main effect of N_{high} on hit rate, $F(1.35, 16.27) = 5.29, p = 0.03$. This reflects the result in 1.2A that it is easier to detect a change the larger that change is. This seems to be a larger effect in the Ellipse experiment, reflected in statistics. There was a non-significant effect of experiment, but a significant interaction between the two, $F(2.03, 24.43), p = 0.01$.

There is an interesting reverse in the qualitative trend when splitting up the hit rate by the reliability of the changed item. There is a significant decrease in hit rate with increasing N_{high} in both low-reliability ($F(1.76, 21.06) = 23.36, p < 0.001$) and high-reliability items ($F(3, 36) = 35.44, p < 0.001$). The intuition behind this decrease is similar to the intuition behind the decrease in false alarms. As the number of high-reliability ellipses increase (and thus high-reliability distractors), the distractors provide more reliable evidence of no change. Thus, the participant is less likely to report change. This trend is reflected in the psychometric function when conditioning on the reliability of the changed stimulus (Figure 1.2C). Additionally, there was a significant effect of experiment on hit rates for high-reliability items only ($F(1, 12) = 14.66. p = 0.002$). It is not clear why the proportion report "change" is lower for the Line experiment; maybe the participant is overcompensating for the change in stimulus type across displays, and thus missing orientation changes.

There are clearly qualitative differences between the data in the two experiments, but it is unclear what that says about the maintenance and use of uncertainty. We use computational modeling to help understand this process.

1.3 Modeling methods

We first replicated the models from Keshvari et al., 2012, then we extended them to be able to fit data from the Line experiment. We model the observer's decision process as consisting of an encoding stage and decision stage. The encoding stage (illustrated in Fig. 1.3) incorporates the task statistics and our assumptions about how the observer generates memories. In the decision stage, the observer calculates a decision variable based on their knowledge of the encoding stage and decides whether to report "change" or "no change" based on some decision rule. We compare models with different formulations of the encoding stage, the observer's assumption about encoding stage, and their decision rule. This section will describe how all models were defined, fit, and compared.

1.3.1 Encoding stage

In this section, we define the statistical structure of the experiment and define our assumptions about how memories are generated in an observer. In Figure 1.3, we display the generative model in turns of variables (left) and an example instantiation of values for a trial (right). The generative model illustrates the statistical dependencies between variables. Each node represents a probabilistic variable, and each arrow indicates a probabalistic dependency. For example, $\mathbf{\Delta}$ depends on Δ.

Variable C indicates whether there was a change on a trial (0: no change, 1: change). The probability of change occurring on each trial is set to 0.5, so $p(C) = 0.5$. If $C = 1$, any item is equally probable to be changed. All the orientations of the items presented on the

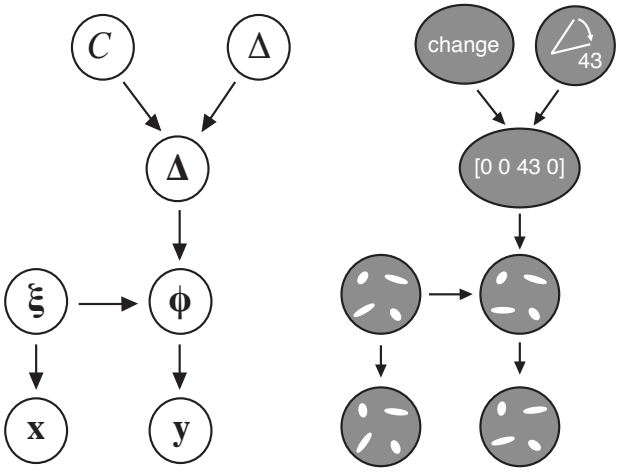

Figure 1.3 Generative model. This illustrates the probablistic dependencies between variables. Each node represents the probabilistic variable, and each arrow indicates a probabalistic dependency. I show the same diagram in terms of the actual variable names (left) and a schematic example (right).

first display, which we denote by vector $\boldsymbol{\xi}^1$, are independently drawn from a uniform distribution over orientation space,

$$p\left(\boldsymbol{\xi}\right) = \left(\frac{1}{2\pi}\right)^N.$$

In all model specifications, we double the actual orientation of stimuli, so the values span 0 to 2π rather than 0 to π. The orientation change, Δ, is drawn from a uniform distribution, $p(\Delta) = \frac{1}{2\pi}$. The vector of changes across all locations, $\boldsymbol{\Delta}$, is a vector of 0s if $C = 0$. If $C = 1$, $\boldsymbol{\Delta}$ is a vector of 0s except for Δ at the location of change i.

$$p(\boldsymbol{\Delta}|C, \Delta) = \frac{1}{N}\sum_{i=1}^{N}\delta(\boldsymbol{\Delta} - C\Delta\mathbf{1}_i),$$

[1] Here, there is a slight difference between Keshvari et al. (2012) and my notation. While Keshvari used $\boldsymbol{\theta}$ as the orientations on the first display, I reserve $\boldsymbol{\theta}$ to represent the vector of parameter values.

where $\mathbf{1}_i$ is a vector of 0s with 1 at the i^{th} entry. The delta function describes a distribution that has zero probability density everywhere except for the point at 0. Thus, this delta function describes a distribution with 0 probability everywhere except when $\boldsymbol{\Delta} = C\Delta\mathbf{1}_i$. Because the change in location is equally probable to occur in each of the N items, we average over the N delta functions describing the change at each location. Note that when $C = 0$, $C\Delta\mathbf{1}_i = 0$, and thus $\boldsymbol{\Delta}$ is a vector of zeros with length N.

The orientations at the second display, $\boldsymbol{\phi}$, is the sum of $\boldsymbol{\xi}$ and $\boldsymbol{\Delta}$,

$$p(\boldsymbol{\phi}|\boldsymbol{\Delta}, \boldsymbol{\xi}) = \delta(\boldsymbol{\phi} - (\boldsymbol{\xi} + \boldsymbol{\Delta})).$$

The noisy measurements of each item on each display, $\boldsymbol{x} = (x_1, ..., x_N)$ and $\boldsymbol{y} = (y_1, ..., y_N)$, respectively, is drawn from a Von Mises distribution centered on the actual orientation presentation,

$$p(\boldsymbol{x}|\boldsymbol{\xi}; \boldsymbol{\kappa_x}) = \prod_{i=1}^{N} p(x_i|\xi_i, \kappa_{x,i}) = \prod_{i=1}^{N} \frac{1}{2\pi I_0(\kappa_{x,i})} e^{\kappa_{x,i}\cos(x_i - \theta_i)}$$
$$p(\boldsymbol{y}|\boldsymbol{\phi}; \boldsymbol{\kappa_y}) = \prod_{i=1}^{N} p(y_i|\phi_i, \kappa_{y,i}) = \prod_{i=1}^{N} \frac{1}{2\pi I_0(\kappa_{y,i})} e^{\kappa_{y,i}\cos(y_i - \phi_i)}.$$

The κs are the concentration parameter of the Von Mises distribution, and are related to the precision with which each item is remembered. The subscript of each κ indicates which item it refers to (e.g., $\kappa_{x,i}$ is for the i^{th} item the first stimulus presentation x_i). We consider Fixed Precision and Variable Precision encoding of items (van den Berg et al., 2012; Fougnie et al., 2012). With a Fixed Precision assumption of encoding noise, the κ for each item is determined only by its ellipse reliability; items with high ellipse reliability would be encoded with parameter κ_{high}, and the lower reliability ellipse with κ_{low}.

With a Variable Precision encoding scheme, $\kappa_{x,i}$ and $\kappa_{y,i}$ are themselves random variables, rather than single values. Rather than sampling κ itself, we sample the Fisher information of the Von Mises distribution, J, from a gamma distribution:

$$p(J) = \frac{1}{\Gamma(k)\tau^k} J^{k-1} e^{J/\tau},$$

where τ is the scale parameter of the gamma distribution, $k = \frac{\bar{J}}{\tau}$, and $\bar{J}$ is the mean precision. The relationship between J and κ is the following:

$$J = \kappa \frac{I_1(\kappa)}{I_0(\kappa)},$$

where I_0 is a modified Bessel function of the first kind of order 0, and I_1 is a modified Bessel function of the first kind of order 1. We assume that the precisions of memories corresponding to low-reliability ellipses are drawn from a gamma distribution with mean $\bar{J}_{\text{low}}$, and those corresponding to high-reliability ellipses are with $\bar{J}_{\text{high}}$. There are certainly differences in the precision with which items in the first and second display are maintained, independent of ellipse reliability. However, the contribution that the first and second displays contribute to the measured change, and calculation of the decision variable (Eq. 1.3), are extremely hard to tease apart in the model. Thus, we use one parameter per reliability and recognize that this estimate will be an aggregate of the precisions of the first and second display.

When modeling the Line experiment, we have an additional parameter, $\bar{J}_{\text{line}}$, which corresponds to the mean precision with which each line on the second display is remembered by the observer. The gamma function from which each line's precision is drawn shares the same scale parameter τ as the distributions from which the ellipse precisions are drawn.

1.3.2 Decoding stage

1.3.2.1 Decision variable

The essence of Bayesian inference is that an observer can compute a posterior, and should if they want to maximize performance. In this case, the observer should calculate the probability of the state of the world (i.e., "change" or "no change") given their observations, $p(C|\boldsymbol{x}, \boldsymbol{y})$, which they can compute using Bayes rule. With a scenario in which there are only two states of the world, it is conventient to combine these into a ratio. Thus, we assume the observer calculates, for each item, the ratio of the likelihood of there being change and the likelihood of there being no change:

$$d = \frac{p(C=1|\boldsymbol{x}, \boldsymbol{y})}{p(C=0|\boldsymbol{x}, \boldsymbol{y}} = \frac{p(\boldsymbol{x}, \boldsymbol{y}|C=1)p(C=1)}{p(\boldsymbol{x}, \boldsymbol{y}|C=0)p(C=0)}. \tag{1.1}$$

Details of the derivation can be found in the Supplementary 1.6.1, but this simplifies to the following expression:

$$d = \frac{p(C=1)}{p(C=0)} \frac{1}{N} \sum_{i=1}^{N} d_i, \tag{1.2}$$

where

$$d_i = \frac{I_0(\kappa_{x,i}) I_0(\kappa_{y,i})}{I_0\left(\sqrt{\kappa_{x,i}^2 + \kappa_{y,i}^2 + 2\kappa_{x,i}\kappa_{y,i}\cos(x_i - y_i)}\right)}, \tag{1.3}$$

I_0 is a modified Bessel function of the first kind of order 0, and the κs are the concentration parameters of the noise distributions for the item indicated in the subscript. Intu-

itively, d_i provides a measure of the evidence of change for the i^{th} item. This value increases with $x_i - y_i$, but weighted by a function of the precisions with which each is remembered. These values are averaged in d, providing the optimal measure of evidence of change of the entire display.

When calculating the decision variable, we consider models with different observer assumptions of the memory generating process, independent of the true generative process. In other words, we allow for there to be "model mismatch" between the true generative process and the believed generative process (Orhan & Jacobs, 2014; Beck et al., 2012; Acerbi et al., 2014).

We consider that the observer may have one of the three assumptions[2]

1. Variable precision (V): mean memory precision varies with ellipse shape, and there is additional noise for each item at each presentation.
2. Fixed precision (F): memory precision varies only with ellipse shape.
3. Same precision (S): memory precision is the same throughout the experiment, and does not vary with ellipse shape or anything else.

1.3.2.2 Decision rule

The observer uses this decision variable to decide whether they believe a change occurred. We consider two decision rules: the optimal (O) and max (M) rules. The Bayes-optimal observer responds "change" whenever the probability of there being a change is greater

[2] There was an additional assumption by Keshvari et al. (2012) that we chose not to model: that participants believe their precision is an average of the precisions of the individual items.

than 0.5. This is equivalent to observer responding "change" if the ratio of the likelihood of there being a change and the likelihood of there being no change (Eq. 1.2) is greater than 1:

$$\frac{p_{\text{change}}}{1-p_{\text{change}}}\frac{1}{N}\sum_{i=1}^{N} d_i > 1, \tag{1.4}$$

where p_{change} is the observer's belief of $p(C = 1)$. Note that participants may have incorrect assumptions about the noise in their memory, but still be acting in accordance with Bayesian decision theory (i.e., still using the correct decision rule), resulting in "imperfectly optimal observers" (Maloney & Zhang, 2010).

This is in contrast to the observer using the max rule, who responds "change" whenever the maximum evidence of change is greater than some criterion, k,

$$\max_i d_i > k. \tag{1.5}$$

This observer is not Bayes-optimal, but may still be using probabilistic computation (i.e., may still be using their uncertainty) in the calculation of d_i.

1.3.3 Model prediction

In this section, we describe how we generate model predictions, given a model and its parameters, $p(r|\boldsymbol{\theta})$. If we knew the observers' memories $\boldsymbol{x}$ and $\boldsymbol{y}$, we could trivially calculate what their response would be. However, because we do not have access to that information, we calculate the probability that they will make a response, given the information we

do have.

Since there are only two responses (and the probability of these responses must add up to 1), we can simply calculate the value of one of them. Arbitrarily, I choose to calculate the probability of the observer responding "change." As experimenters, we do not know $\boldsymbol{x}$ and $\boldsymbol{y}$ and thus must marginalize over them:

$$p(r|\boldsymbol{\theta}) = \iint p(r|\boldsymbol{x}, \boldsymbol{y}, \theta) p(\boldsymbol{x}|\boldsymbol{\xi}, \boldsymbol{J_x}) p(\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{J_y}) d\boldsymbol{x} d\boldsymbol{y},$$

where $\boldsymbol{J_x}$ and $\boldsymbol{J_y}$ are the vectors of the precisions of the distributions from which each element in $\boldsymbol{x}$ and $\boldsymbol{y}$ are respectively drawn from. These values are defined by the encoding model. For example, in the Fixed Precision encoding, $J_{x,i} = J_{y,i}$ would be equivalent to J_{low} or J_{high}, depending on the reliability of the ellipse. For the VP model however, $J_{x,i}$ and $J_{y,i}$ are themselves random variables and thus require additional marginalizations:

$$p(r|\boldsymbol{\theta}) = \iiiint p(r|\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}) p(\boldsymbol{x}|\boldsymbol{\xi}, \boldsymbol{J_x}) p(\boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{J_y}) p(\boldsymbol{J_x}|\bar{\boldsymbol{J}}_{\boldsymbol{x}}, \tau) p(\boldsymbol{J_y}|\bar{\boldsymbol{J}}_{\boldsymbol{y}}, \tau) d\boldsymbol{x} d\boldsymbol{y} d\boldsymbol{J_x} d\boldsymbol{J_y}.$$

We numerically integrate over these distributions by sampling. In the following results, we sampled 50 samples for each trial[3].

[3] This may seem like a somewhat low sample number (and it is), but we have fit data with 200, 500, and 1000 samples and have not seen a difference in model fits or model comparison results.

1.3.4 Parameters and parameter estimation

1.3.4.1 Parameters

There are two possible encoding schemes ((V)ariable, (F)ixed), three possible observer assumptions of noise ((V)ariable, (F)ixed, (S)ame), and two possible decision rules ((O)ptimal, (M)ax). Factorially combining each of these characteristics would yield 12 different models. We choose not to consider the two models in which the generative model is F but the observer assumes V under the assumption that people tend not to assume the world is more complicated than it actually is; thus, we test a total of 10 models. We denote each model by the letters corresponding to their encoding scheme, assumption of encoding, and decision rule (e.g., VVO is the model with Variable precision encoding, an observer that assumes Variable precision encoding, and an Optimal decision rule).

Observers with a V encoding scheme have parameters $\bar{J}_{\text{high}}$ and $\bar{J}_{\text{low}}$ corresponding to the mean precision of the high and low reliability ellipses, respectively. The scale parameter, τ, of the gamma distribution from which item-wise precision is drawn is shared across the two ellipse values. If the observer is or believes they are F, then the according precision is $J_{\text{high}} = \bar{J}_{\text{high}}$ and $J_{\text{low}} = \bar{J}_{\text{low}}$ for high and low reliability ellipses, respectively. If the true generative model and observer assumption are both F, then the model does not have the τ parameter.

When fitting data from the Line experiment, there is an additional parameter $\bar{J}_{\text{line}}$. If the observer is or believes they are F, then the according precision is $J_{\text{line}} = \bar{J}_{\text{line}}$. Scale parameter τ is then shared across all three stimulus types.

If the participant has the incorrect assumption that their precision is equal across reliabil-

ities, items, and trials, then there is an additional parameter $J_{\text{assumed, ellipse}}$, corresponding to the assumed precision of all ellipses. When fitting data from the Line experiment, we additionally have parameters $J_{\text{assumed, line}}$, allowing the assumed precisions to differ across shapes.

There is one additional parameter for the decision process. If the observer uses the optimal decision rule, there is parameter p_{change} corresponding to the observer's belief of the prior. While this value is in reality 0.5, we allow the observer to have an incorrect belief. If the observer uses the max rule, then we have parameter k, corresponding to the decision criterion. If any item has a decision variable greater than k, then they will respond "change." Each model and their corresponding parameters are also indicated in Table 1.1.

Model	Parameters
VVO	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, p_{\text{change}}(, \bar{J}_{\text{line}})$
VFO	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, p_{\text{change}}(, \bar{J}_{\text{line}})$
VSO	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, p_{\text{change}}, J_{\text{assumed, ellipse}}(, \bar{J}_{\text{line}}, J_{\text{assumed, line}})$
VVM	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, k(, \bar{J}_{\text{line}})$
VFM	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, k(, \bar{J}_{\text{line}})$
VSM	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, k(, \bar{J}_{\text{line}})$
FFO	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, p_{\text{change}}(, \bar{J}_{\text{line}})$
FSO	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, p_{\text{change}}, J_{\text{assumed, ellipse}}(, \bar{J}_{\text{line}}, J_{\text{assumed, line}})$
FFM	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, k(, \bar{J}_{\text{line}})$
FSM	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, k(, \bar{J}_{\text{line}})$

Table 1.1 Model names and parameters. Model names and parameters for Ellipse condition. Additional Line condition parameters are displayed in parantheses. The first digit of the model corresponds to the encoding scheme: (V)ariable or (F)ixed. The second digit corresponds to the assumed encoding scheme: (V)ariable, (F)ixed, or (S)ame. The third digit corresponds to the decision rule: (O)ptimal or (M)ax.

1.3.4.2 Maximum-likelihood estimation of parameters

The likelihood of the parameter combination ($\boldsymbol{\theta}$) for a given participant is the probability of the data given the parameter combination. We calculate the log likelihood, which we denote LL:

$$\begin{aligned} \mathrm{LL}(\boldsymbol{\theta}) &= \log p(\boldsymbol{\theta}|\text{data, model}) \\ &= \log \prod_{t}^{N_{\text{trials}}} p(r_t|\boldsymbol{\theta}) \\ &= \sum_{\text{“change” trials } j} \log p(r_j = 1|\boldsymbol{\theta}) + \sum_{\text{“no change” trials } j} \log p(r_j = 0|\boldsymbol{\theta}) \end{aligned}$$

Because we are sampling to approximate this value, the calculation of our LL is stochastic and can be computationally expensive. We thus used the optimization algorithm Bayesian Adaptive Direct Search (BADS; Acerbi & Ma, 2017) in MATLAB, which combines two different optimization methods: mesh adaptive direct search and Bayesian optimization. BADS is a more suitable optimization method for stochastic likelihood landscapes (due to sampling) than common optimization functions like MATLAB's fmincon or fminsearch because it explicitly incorporates uncertainty in the estimated LL. Additionally, it is able to converge in fewer function evaluations than stochastic optimization methods like cmaes (Hanson, Niederberger, Guzella, & Koumoutsakos, 2008) and genetic algorithm (Goldberg, 1988). We use 20 different starting positions, using latin hypercube sampling, to minimize the probability of finding a local minimum. We took the minimum negative log-likelihood of all the runs as our estimate of the maximum-likelihood, and the corresponding parameter combination as our ML parameter estimate. We denote the maximum log-likelihood as LL*.

1.3.5 Model comparison

Models with more parameters are more flexible, usually providing a better fit to the data. To encourage parsimony, we compared models using corrected Akaike Information Criterion (AICc; Hurvich & Tsai, 1987) and the Bayesian Information Criterion (BIC; (Schwarz, 1978)). BIC penalizes slightly harsher than AICc.

$$\text{AICc} = -2\text{LL}^* + 2n_{\text{pars}} + \frac{2n_{\text{pars}} + 1}{n_{\text{trials}} - n_{\text{pars}} - 1}$$

$$\text{BIC} = -2\text{LL}^* + 2n_{\text{pars}} \log n_{\text{trials}}$$

1.4 Results

In this section, we look at the results of the model fits. We fit 10 models to just the Ellipse experiment, with the goal of replicating the findings of Keshvari et al., 2012. We also collected the Line experiment to test if people maintained uncertainty during the delay. To test this, we fit the same 10 models to the data from the Line experiment alone as well as jointly fit the data from the Ellipse and Line experiment. We chose to jointly fit both experiment's data because jointly fitting theoretically allows us to get a more reliable estimate of the precision with which low- and high-reliability ellipses are remembered. However, we do not allow the strategies to differ across the Ellipse and Line experiments when jointly fitting. If people do not use the same strategy across experiments, then the results of jointly fitting will not be meaningful. Thus, we also separately fit the data from the Line experiment, which provided a better estimate of which strategy each participant was using in that experiment. We investigate each model's absolute goodness-of-fit to the data

through visualization (e.g., Figure 1.4) and their relative goodness-of-fit through AICc and BIC (Figure 1.6).

Fitting Ellipse data only. When fitting just Ellipse data, we find the best performing model is the VVO model, which defines the observer to have Variable precision encoding, a Variable precision assumed encoding, and the optimal decision rule. The only model which performs better for a subset of participants is the VVM model, which has the same encoding and assumption of encoding, but a different decision rule. The number of participants best fit by these two models is split, the VVO model fitting better for seven participants and the VVM model fitting better for the remaining six. The VVM model fits better for most participants for most models, except one participant for the VFO, one for the FFO, and the aforementioned seven for the VVO model. The two model comparison metrics gave consistent results, so just BIC differences between each model and the VVO model are illustrated in the left plot of Figure 1.6. Additionally, we can see in the first row of Figure 1.4 that the VVO (and VVM) models provide a good qualitative fit to the data. Parameter estimates for the VVO and VVM models are in Supplementary (Tables 1.2 and 1.3, respectively).

Fitting Line data only. Model predictions are illustrated in Figure 1.5. The differences in model predictions are less obvious than the fits of the Ellipse data because there are less differences between high- and low-reliability (Figure 1.6). Nevertheless, we find the best performing model when fitting just Line data is the VVO model, which defines the observer to have Variable precision encoding, a Variable precision assumed encoding, and the optimal decision rule. There is one participant who is better fit by a model that defines an observer who ignores uncertainty (assumes Same precision regardless of stimulus reliability) and uses the optimal decision rule, regardless of whether their memories are truly

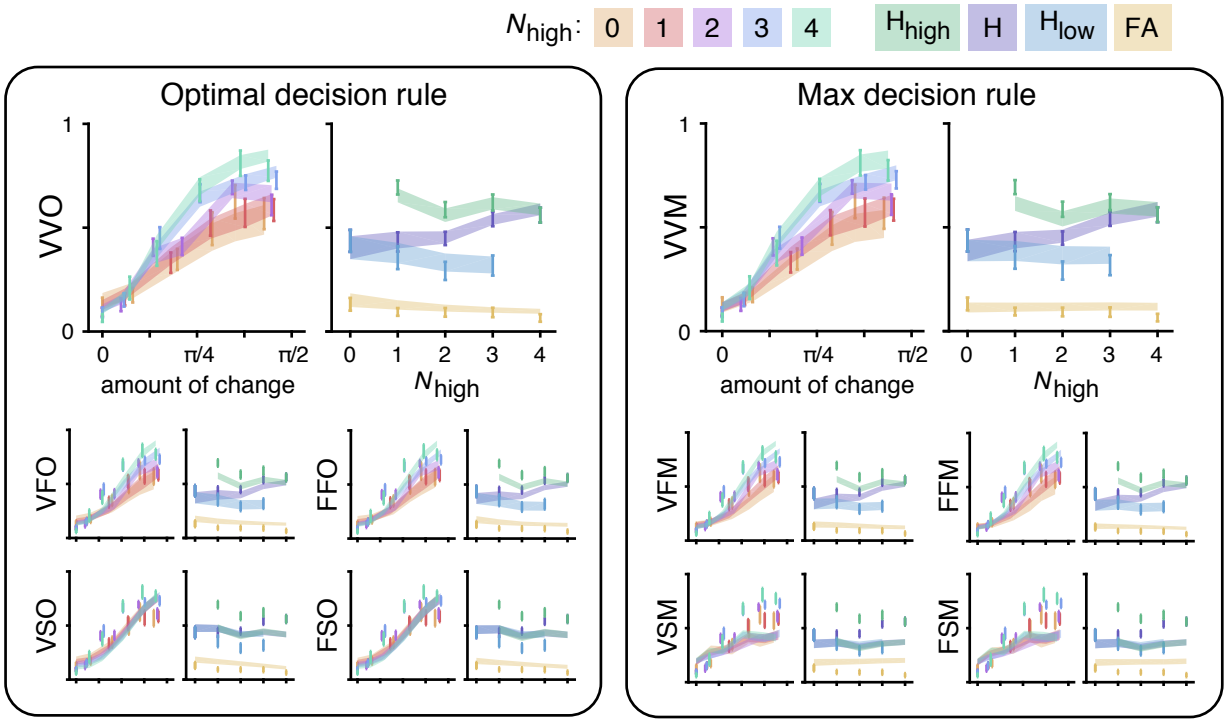

Figure 1.4 Ellipse experiment model fits. $M \pm SEM$ data (error bars) and model predictions (fills) for all models, organized by decision rule. For each model, the left graph illustrates the proportion report "change" as a function of amount of change. Color indicates the number of high reliability ellipses. The right graph illustrates the proportion hits for high-reliability items (green), hits for low-reliability items (blue), hits averaged across the display (purple), and false alarms (gold) as a function of number of high-reliability items.

encoded with Variable of Fixed precision. Parameter estimates for the VVO model is in Supplementary (Table 1.4).

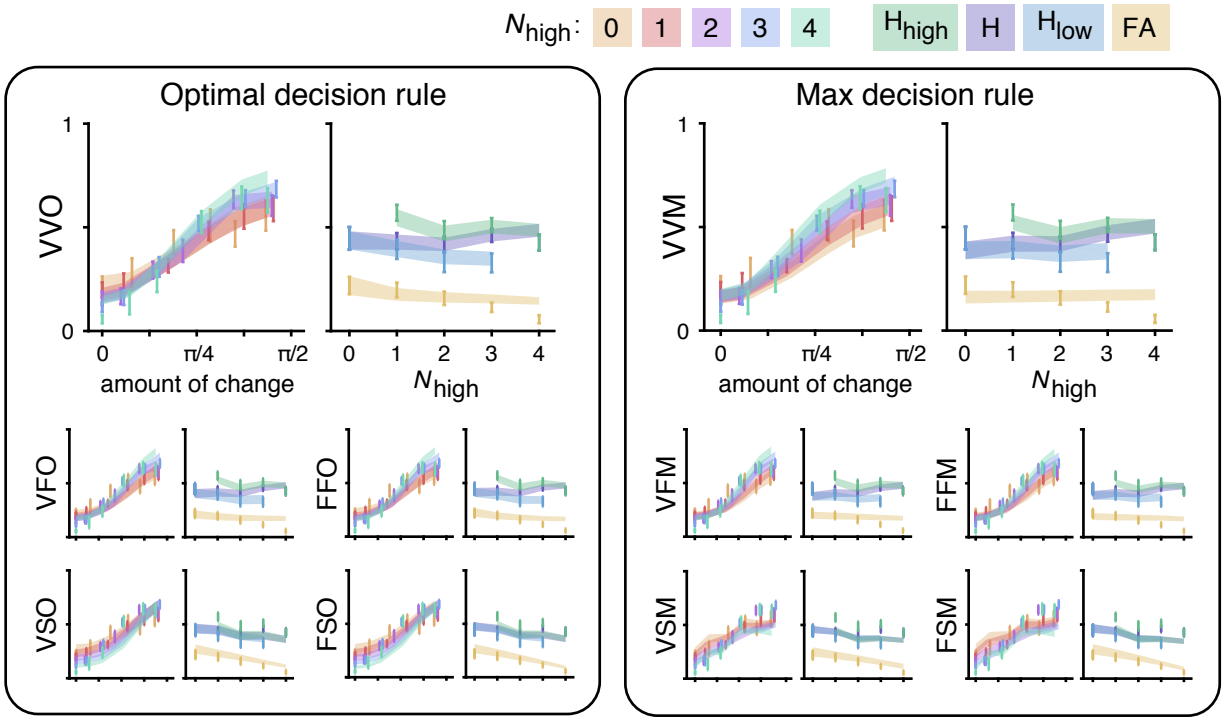

Figure 1.5 Line experiment model fits. $M \pm SEM$ data (error bars) and model predictions (fills) for all models, organized by decision rule. For each model, the left graph illustrates the proportion report "change" as a function of amount of change. Color indicates the number of high reliability ellipses. The right graph illustrates the proportion hits for high-reliability items (green), hits for low-reliability items (blue), hits averaged across the display (purple), and false alarms (gold) as a function of number of high-reliability items.

Jointly fitting Ellipse and Line data. Unsurprisingly, when jointly fitting the data, the results of the model comparison are in between the results when fitting the experiments separately. The VVO model fits best for all participants for most models, and for 9 of 13 participants for the VVM model. Again, this model provides a good qualitative fit to the data (top row of Fig 1.7). The VVM model fits are qualitatively reasonable (second row of Fig 1.7), but do show more deviations in the data when compared to the VVO

model. Quantitatively, it does not perform as well as the VVO model, fitting worse than for several participants across different models. Parameter estimates for the VVO and VVM models are in Supplementary (Tables 1.5 and 1.6, respectively).

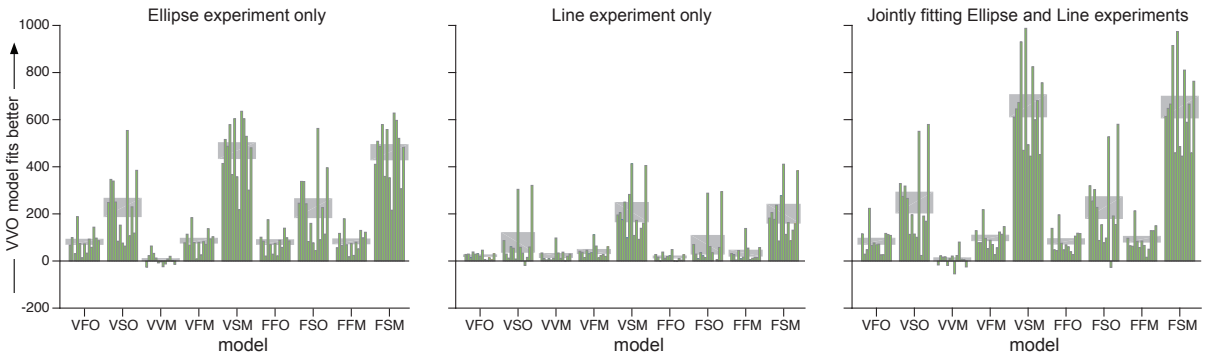

Figure 1.6 Model comparison. BIC(VVO)-BIC(model) for fitting data from the Ellipse experiment (*left*), Line experiment (*middle*), or both experiments (*right*). Individual participant shown in bars; $M \pm SEM$ shown in grey. Greater value indicates worse fit in comparison to the VVO model.

1.5 Discussion

In this chapter, we investigated how uncertainty was maintained and used in working memory. Specifically, we hypothesized that uncertainty would be accurately maintained during a delay and used when making an orientation change detection decision. We addressed this question by first replicating the results of Keshvari et al. (2012), then adjusted it slightly by changing the ellipses in the second presentation to lines.

1.5.1 Summary of results

First, we successfully replicated the results of Keshvari and others. We found that people were best fit by a model with Variable precision encoding and an observer knows their un-

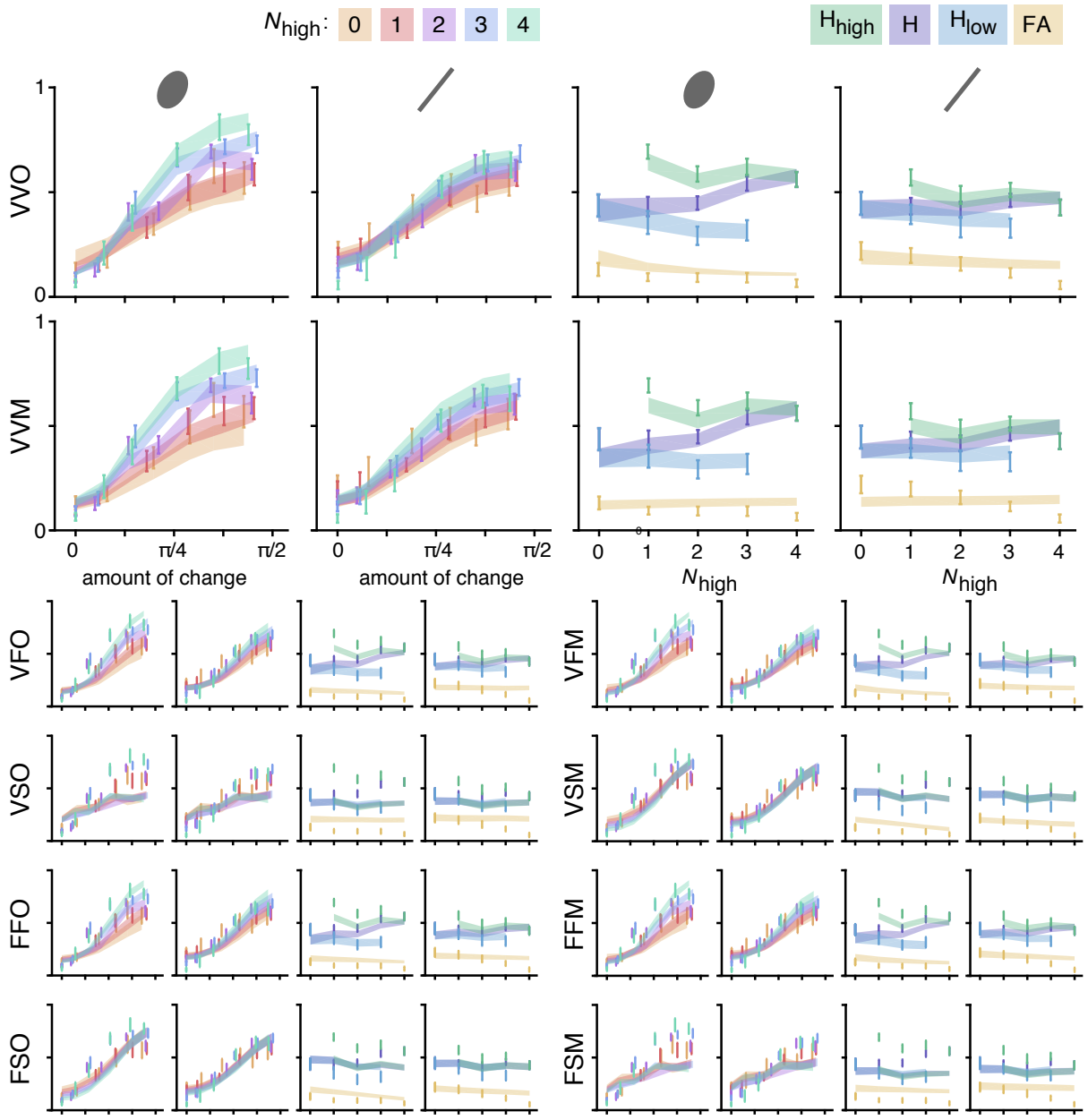

Figure 1.7 Joint model fits. $M \pm SEM$ of data (error bars) and model predictions (fills) for all data and models. Each column corresponds to data , as illustrated in 1.2: psychometric functions for the Ellipse experiment, psychometric function for the Line experiment, hits and false alarms for the Ellipse experiment, and hits and false alarms for the Line experiment. The two winning models are shown as the first two rows, and all other models shown below.

certainty. We found that seven of 13 participants were best fit by a model in which the observer has the optimal decision rule; the other six were best fit by a model with a max decision rule. At first glance, this seems like a departure from the results of Keshvari et al., who claim to have found a "decisive difference" between the VVO and VVM model. However, I would argue that a $M \pm SEM$ difference of 15.4 ± 17.3 between the models is hardly decisive at all. Indeed, I would conclude that, in both studies, there is no clear evidence of optimality in either datasets, just probabilistic computation. These results are consistent with previous findings that, in a variety of perceptual decision-making tasks including change detection, the optimal decision rule almost always fits data better than the max rule, except for the odd case in which both rules perform similarly well. (Ma, Shen, Dziugaite, & van den Berg, 2015).

There were a couple methodological differences between how we fit and compared models. Keshvari and others compared models using marginal likelihoods while we used maximum-likelihoods; we calculated the $p(r|\boldsymbol{\theta})$ in subtly different ways. It is comforting that despite these methodological differences, we found remarkably consistent results. This is strong evidence that people are indeed using item-specific uncertainty in a four-item orientation change detection task.

We found consistent results when separately or jointly fitting the data from the Line experiment with that of the Ellipse experiment, perhaps with larger support for optimality, in which uncertainty would have to be maintained to be used. Again, we found that participants were best fit by a model that had Variable precision encoding and an observer's assumption of Variable precision encoding. When solely fitting Line data, we found that all but one participant was best fit by the VVO model. When jointly fitting data from both experiments, 9 of the 13 participants were best fit by the VVO model, max decision

rule fit best for the remaining 4. Unlike the Ellipse experiment, these results suggest that information was actually *maintained* in working memory, since the information was not available to them through some the ellipse reliability.

1.5.2 Limitations and future directions

In this chapter, we were somewhat unable to distinguish optimal and max decision rules. These decision rules are notoriously hard to tease apart (Ma et al., 2015), but there may be experimental choices that may better distinguish these two decision rules. In this task, the optimal decision rule (Eq. 1.4 is much like a weighted average) and the max decision rule is, as the name implies, a max (Eq. 1.5). In our current design, $p(C = 1) = 0.5$ and, in change trials, each item is equally likely to change. Intuitively, the optimal and max rules make the same predictions in this task because the aspects that only the optimal decision rule considers do not change the predicted responses, relative to the max rule. If we biased $p(C = 1)$ or the probability of each item changing given a change trial, the max and optimal decision rules may result in different predictions. This is of course something we can test through model simulations. Distinguishing between these two decision rules, rules that in most case provide similar predictions, may provide some insight into whether humans are actually performing Bayesian inference or some type of approximate inference that provides near-optimal performance in most cases.

The results of the parameter estimates in the Line experiment are a potential limitation of the current results. Earlier, we pointed out that the qualitative lack of effect of reliability on proportion report "change" may suggest that uncertainty is not being used. However, the VVO model, a model in which the observer maintains and uses uncertainty optimally, fit the data better than all other models. This model captures this seeming lack of effect

of reliability by estimating the parameter of $\bar{J}_{\text{line}}$, the average precision for the line stimuli, to be about as low as $\bar{J}_{\text{low}}$, the average precision for the low-reliability ellipse (see Table 1.4). This is counterintuitive; the precision with which participants represent the line stimulus should be at a higher precision than either of the ellipses because it has the highest-reliability orientation information. In other words, $\bar{J}_{\text{line}}$ should be greater than $\bar{J}_{\text{high}}$ and $\bar{J}_{\text{low}}$. This is clearly not true. This result may seem to completely invalidate the results of the Line experiment. However, the change of stimulus in the Line experiment (going from ellipse to line) does affect the task subjectively, which isn't necessarily accounted for in the model. Perhaps the change in stimulus results in a less precise representation of orientation information overall, which is somehow reflected solely in the lower $\bar{J}_{\text{line}}$ estimates. Alternatively, people may not be maintaining and using uncertainty, but somehow the VVO model was able to capture the data better than the other tested models. This is an aspect of the result that we do not fully understand and could significantly affect the interpretation of the results. We are investigating this in more detail.

If the modeling results, however, truly reflect a maintenance of uncertainty, it is of course still possible that the observer is maintaining some heuristic, not uncertainty per se. For example, participants could maintain the ellipse reliability, using the knowledge that the sensory reliability modulates uncertainty rather than representing a probability distribution over the stimulus value (Barthelme & Mamassian, 2010). Unfortunately, our task cannot separate these two possibilities. Additionally, it is always possible that people are simply learning a stimulus-response mapping rather than performing probabilistic computations or Bayesian inference (Maloney & Mamassian, 2009). We attempted the minimize the possibility of this when designing the experiment, by choosing to withhold feedback throughout the experiment.

1.5.3 Conclusions

In this chapter, we investigated and found that people maintain and use an accurate representation of uncertainty when making change detection decisions. While the results of this study are potentially promising, more work needs to be done to understand the role of uncertainty in a broad range of working memory tasks. In Chapter 3, we attempt to generalize these results by investigating the use of uncertainty in an entirely different task. Importantly, we incorporate one key aspect of Bayesian decision models that we ignore in this chapter: a cost function.

1.6 Supplementary

1.6.1 Derivation of decision variable

In this section, we compute the decision variable from Eq. 1.1. In order to compute this value, we must marginalize over the unknown variables $\boldsymbol{\xi}, \boldsymbol{\phi}, \Delta$, and Δ.

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}|C)p(C) &= \iiiint p(\boldsymbol{x}|\boldsymbol{\xi})p\left(\boldsymbol{\xi}\right) p(\boldsymbol{y}|\boldsymbol{\phi})p(\boldsymbol{\phi}|\boldsymbol{\Delta})p(\boldsymbol{\Delta}|C,\Delta)p(\Delta)p(C)d\boldsymbol{\xi}d\boldsymbol{\phi}d\boldsymbol{\Delta}d\Delta \\ &= p(C)\left(\frac{1}{2\pi}\right)^{N+1} \iiiint p(\boldsymbol{x}|\boldsymbol{\xi})p(\boldsymbol{y}|\boldsymbol{\phi})\delta(\boldsymbol{\phi} - (\boldsymbol{\xi} + C\boldsymbol{\Delta})) \\ &\quad \left(\frac{1}{N}\sum_{i=1}^{N}\delta(\boldsymbol{\Delta}_i - C\mathbf{1}_i)\right) d\boldsymbol{\xi}d\boldsymbol{\phi}d\boldsymbol{\Delta}d\Delta \\ &= p(C)\left(\frac{1}{2\pi}\right)^{N+1} \frac{1}{N}\sum_{i=1}^{N}\iint p(\boldsymbol{x}|\boldsymbol{\xi})p(\boldsymbol{y}|\boldsymbol{\xi} + C\Delta\mathbf{1}_i)d\boldsymbol{\xi}d\Delta \end{aligned}$$

Then, plugging this equation into the likelihood ratio:

$$\begin{aligned} \frac{p(\mathbf{x}, \mathbf{y}|C=1)p(C=1)}{p(\mathbf{x}, \mathbf{y}|C=0)p(C=0)} &= \frac{p(C=1)}{p(C=0)} \frac{\sum_{i=1}^{N}\iint p(\boldsymbol{x}|\boldsymbol{\xi})p(\boldsymbol{y}|\boldsymbol{\xi} + C\Delta\mathbf{1}_i)d\boldsymbol{\xi}d\Delta}{\sum_{i=1}^{N}\iint p(\boldsymbol{x}|\boldsymbol{\xi})p(\boldsymbol{y}|\boldsymbol{\xi})d\boldsymbol{\xi}d\Delta} \\ &= \frac{p(C=1)}{p(C=0)} \frac{\sum_{i=1}^{N}\iint p(\boldsymbol{x}|\boldsymbol{\xi})p(\boldsymbol{y}|\boldsymbol{\xi} + C\Delta\mathbf{1}_i)d\boldsymbol{\xi}d\Delta}{2\pi N\int p(\boldsymbol{x}|\boldsymbol{\xi})p(\boldsymbol{y}|\boldsymbol{\xi})d\boldsymbol{\xi}} \end{aligned}$$

Because the N items are conditionally independent, we can break the expression up into a product of each item and further simplify.

$$
\begin{aligned}
d &= \frac{p(C=1)}{p(C=0)} \frac{\sum_{i=1}^{N} \int \left(\prod_{j \neq i} \int p(x_j|\xi_j) p(y_j|\xi_j) d\xi_j \right) \left(\int p(x_i|\xi_i) p(y_i|\xi_i + \Delta) d\xi_i \right) d\Delta}{2\pi N \prod_{i=1}^{N} \int p(x_i|\xi_i) p(y_i|\xi_i) d\xi_i} \\
&= \frac{p(C=1)}{p(C=0)} \sum_{i=1}^{N} \frac{\int \left(\prod_{j \neq i} \int p(x_j|\xi_j) p(y_j|\xi_j) d\xi_j \right) \left(\int p(x_i|\xi_i) p(y_i|\xi_i + \Delta) d\xi_i \right) d\Delta}{2\pi N \left(\prod_{j \neq i} \int p(x_j|\xi_j) p(y_j|\xi_j) d\xi_j \right) \left(\int p(x_i|\xi_i) p(y_i|\xi_i) d\xi_i \right)} \\
&= \frac{p(C=1)}{p(C=0)} \sum_{i=1}^{N} \frac{\iint p(x_i|\xi_i) p(y_i|\xi_i + \Delta) d\xi_i d\Delta}{2\pi N \int p(x_i|\xi_i) p(y_i|\xi_i) d\xi_i} \\
&= \frac{p(C=1)}{p(C=0)} \sum_{i=1}^{N} \frac{\iint \mathrm{VM}(x_i; \xi_i, \kappa_{x,i}) \mathrm{VM}(y_i; \xi_i + \Delta, \kappa_{y,i}) d\xi_i d\Delta}{2\pi N \int \mathrm{VM}(x_i; \xi_i, \kappa_i) \mathrm{VM}(y_i; \xi_i, \kappa_i) d\xi_i} \\
&= \frac{p(C=1)}{p(C=0)} \sum_{i=1}^{N} \frac{\iint \mathrm{VM}(\xi_i; x_i, \kappa_i) \mathrm{VM}(\Delta; y_i - \xi_i, \kappa_{y,i}) d\xi_i d\Delta}{2\pi N \int \mathrm{VM}(x_i; \xi_i, \kappa_{x,i}) \mathrm{VM}(y_i; \xi_i, \kappa_{y,i}) d\xi_i} \\
&= \frac{p(C=1)}{p(C=0)} \sum_{i=1}^{N} \frac{1}{2\pi N \int \mathrm{VM}(x_i; \xi_i, \kappa_{x,i}) \mathrm{VM}(y_i; \xi_i, \kappa_{y,i}) d\xi_i} \\
&= \frac{p(C=1)}{p(C=0)} \sum_{i=1}^{N} \frac{1}{2\pi N \int \frac{I_0(\kappa)}{2\pi I_0(\kappa_{x,i}) I_0(\kappa_{y,i})} \mathrm{VM}\left(\xi_i; \mu, \kappa\right) d\xi_i},
\end{aligned}
$$

where $\mu = x_i + \arctan(\sin(y_i - x_i), (\kappa_{x,i}/\kappa_{y_i}) + \cos(y_i - x_i))$ and $\kappa = \sqrt{\kappa_{x,i}^2 + \kappa_{y,i}^2 + 2\kappa_{x,i}\kappa_{y,i}\cos(x_i - y_i)}$.

$$
\begin{aligned}
&= \frac{p(C=1)}{p(C=0)} \sum_{i=1}^{N} \frac{I_0(\kappa_{x,i}) I_0(\kappa_{y,i})}{N I_0\left(\kappa\right) \int \mathrm{VM}\left(\xi_i; \mu, \kappa\right) d\xi_i} \\
&= \frac{p(C=1)}{p(C=0)} \frac{1}{N} \sum_{i=1}^{N} \frac{I_0(\kappa_{x,i}) I_0(\kappa_{y,i})}{I_0\left(\kappa\right)},
\end{aligned}
$$

This produces the final expression of the decision variable. Eq. 1.2,

$$
d = \frac{p(C=1)}{p(C=0)} \frac{1}{N} \sum_{i=1}^{N} d_i.
$$

where

$$d_i = \frac{I_0(\kappa_{x,i})I_0(\kappa_{y,i})}{I_0\left(\sqrt{\kappa_{x,i}^2 + \kappa_{y,i}^2 + 2\kappa_{x,i}\kappa_{y,i}\cos(x_i - y_i)}\right)}.$$

1.6.2 Maximum-likelihood parameter estimates

I report summary statistics for only the models that provided reasonable qualitative fits to the data.

	$\bar{J}_{\text{high}}$	$\bar{J}_{\text{low}}$	τ	p_{change}
mean	28.06	10.84	43.74	0.46
SEM	5.05	2.20	10.85	0.02

Table 1.2 VVO parameter estimates when fitting Ellipse experiment. Mean and SEM across participants for all parameters in the model in which participants have variable precision (VP) encoding, assume VP encoding, and use an optimal decision rule.

	$\bar{J}_{\text{high}}$	$\bar{J}_{\text{low}}$	τ	k
mean	22.32	11.85	33.04	0.84
SEM	2.40	2.11	5.59	1.02

Table 1.3 VVM parameter estimates when fitting Ellipse experiment. Mean and SEM across participants for all parameters in the model in which participants have variable precision (VP) encoding, assume VP encoding, and use a max decision rule.

	$\bar{J}_{\text{high}}$	$\bar{J}_{\text{low}}$	$\bar{J}_{\text{line}}$	τ	p_{change}
mean	22.51	5.62	5.62	5.78	0.45
SEM	4.65	2.35	1.33	1.11	0.02

Table 1.4 VVO parameter estimates when fitting Line experiment. Mean and SEM across participants for all parameters in the model in which participants have variable precision (VP) encoding, assume VP encoding, and use an optimal decision rule.

	$\bar{J}_{\text{high}}$	$\bar{J}_{\text{low}}$	$\bar{J}_{\text{line}}$	τ	p_{change}
mean	22.16	9.52	9.69	31.44	0.47
SEM	4.36	2.08	1.61	8.66	0.02

Table 1.5 VVO parameter estimates when jointly fitting Ellipse and Line experiment. Mean and SEM across participants for all parameters in the model in which participants have variable precision (VP) encoding, assume VP encoding, and use an optimal decision rule.

	$\bar{J}_{\text{high}}$	$\bar{J}_{\text{low}}$	$\bar{J}_{\text{line}}$	τ	k
mean	16.25	7.32	10.23	23.13	0.05
SEM	2.25	1.57	1.81	6.83	0.85

Table 1.6 VVM parameter estimates when jointly fitting Ellipse and fitting Line experiment. Mean and SEM across participants for all parameters in the model in which participants have variable precision (VP) encoding, assume VP encoding, and use a max decision rule.

2 Strategic allocation of working memory resource to minimize behavioral loss

It ain't no fun if the homies can't have none.

Snoop Dogg

2.1 Introduction

In the previous chapter, the precision with which an item was encoded was manipulated through the ellipse reliability of the item. However, in real life, the mean precision of memory for items does not solely depend on bottom-up characteristics of the item. The paradigm used in the previous chapter, along with a bulk of the working memory literature, ignores an important aspect of the real world: not all items are equally relevant. In this chapter, we investigate how behavioral relevance, or priority, affects our working memory representations.

Studies that have investigated the effects of priority on behavior typically involve an attentional cue that indicates the priority of items, in which one item is more likely to be probed or rewarded higher than the rest of the items in the display. Perhaps intuitively, these results show that people are able to remember a more important item more precisely than less important items(Bays, 2014; Dube et al., 2017; Emrich et al., 2017; Klyszejko et al., 2014; Gorgoraptis et al., 2011).

Importantly, previous studies have only used two levels of priority: one high priority item (the attended or valid item) and low-priority items (the unattended or invalid targets). These results could theoretically be explained by something like the spotlight or focus of attention (Posner, Snyder, & Davidson, 1980; Treisman & Sato, 1990; Oberauer, 2002), in which one item or area in space is afforded special privilege and thus has higher memory precision.

However, we believe memory is more flexible than that; we believe people have some limited amount of resource that can be flexibly allocate across items. There are a few compu-

tational models that hypothesize different allocation strategies, and each has fit the data relatively well. Emrich et al. (2017) proposed that resource is allocated in approximate proportion to the probe probabilities. Bays (2014) proposed that resource is allocated such that the expected squared error is minimized. Sims (2015) proposed more generally that resource is allocated to minimize loss. However, these models have conflicting hypotheses.

To distinguish these models, we designed a task that had three nonzero levels of priority on every trial. We used a four-item delayed-estimation task, in which participants remembered the location of items in space, and estimated the location of a probed item with a memory guided saccade. Importantly, the task had an attentional precue indicating the priority of each item, which was operationalized as probe probability. We predicted that people would allocate resource according to priority, and we fit and compared different models to determine what strategy people use when allocating their resource. We found that people are best described by a model that assumes people allocate resource in order to minimize behavioral loss.

2.2 Experimental methods

2.2.1 Participants

Fourteen participants (5 males, mean age=30.3, SD=7.2) participated in this experiment. Everyone had normal or corrected-to-normal vision and no history of neurological disorders. Participants were naive to the study hypotheses and were paid $10/hour. We obtained informed, written consent from all participants. The study was in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of New

York University.

2.2.2 Apparatus

Participants were placed 56 cm from the monitor (19 inches, 60 Hz), with their heads in a chinrest. Eye movements were calibrated using the 9-point calibration and recorded at a frequency of 1000 Hz (Eyelink 1000, SR Research). Target stimuli were programmed in MATLAB (MathWorks) using the MGL toolbox (Gardner Lab, Stanford) and were displayed against a uniform grey background.

2.2.3 Trial procedure

Each trial (Fig. 2.1) began with a 300 ms increase in the size of the fixation symbol, an encircled fixation cross. This was followed by a 400 ms endogenous precue, consisting of three colored wedges presented within the fixation symbol, each of which angularly filled one quadrant. The radial sizes and colors of the wedges corresponded to probe probabilities of 0.6, 0.3, and 0.1, respectively. (In our illustrations, the priorities correspond to colors orange, yellow, and green, respectively, but in the experiment the wedges were pink, yellow, and blue, respectively.) The quadrant with a probe probability of 0.0 did not have a wedge.

The precue was presented for 400 ms; this was followed by a 700 ms interstimulus interval, then by the targets, presented for 100 ms. The targets were four dots, each in separate visual quadrants. The dots were presented at approximately 10 degrees of visual angle from fixation, with random jitter of 1 degree of visual angle to each location. The location of the targets in polar coordinates were pseudo-randomly sampled from every 10 degrees,

avoiding cardinal axes.

This was followed by a variable delay, chosen with equal probability from the range between 1000 and 4000 ms in 500 ms increments. A response cue appeared afterward, which was a white wedge that filled an entire quadrant of the fixation symbol. Participants were instructed to make a saccade to the remembered dot location within the corresponding quadrant of the screen.

After the saccade, the actual dot location was presented as feedback and the participant made a corrective saccade to that location. After 500 ms, the feedback disappeared, participants returned their gaze to the central fixation cross, and a 1500 ms inter-trial interval began. Participants completed between 10 to 15 forty-trial runs over one or two one-hour sessions.

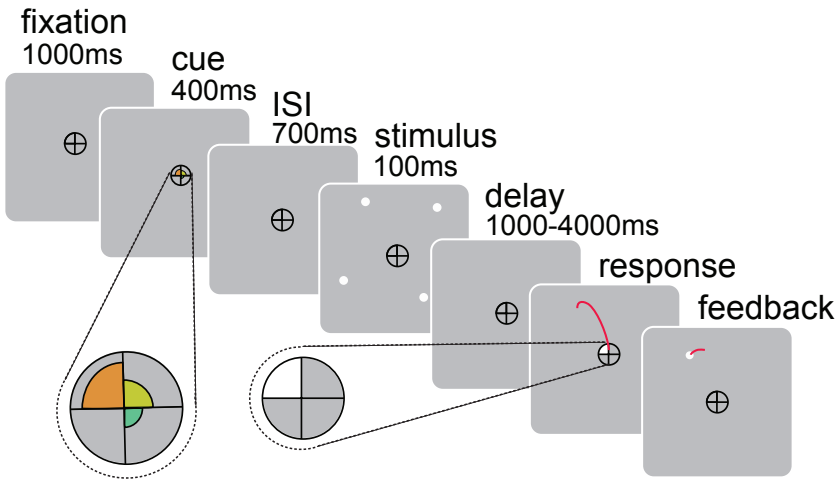

Figure 2.1 Trial sequence. Participants remembered the location of four dots (with different priorities) over a variable delay, were probed to respond to the location of one, made a memory-guided saccade to the remembered location of the probed item, and made a corrective saccade to the actual location of the item.

2.2.4 Data processing

Processing and manual scoring of eye movement data were performed in an in-house MATLAB function-graphing toolbox (iEye). Eye position and saccadic reaction time (SRT) were extracted from iEye. We excluded trials in which a) participants were not fixating in the middle of the screen during stimulus presentation, b) saccades were initiated before 100 ms or after 1200 ms after the response cue onset, c) pupil data during the response period were missing, or d) participants made a saccade to the wrong quadrant, ignoring the response cue. This resulted in removing between 1% and 7% of trials per subject.

2.3 Modeling methods

2.3.1 Encoding stage

We model target location $\boldsymbol{s}$ as a two-dimensional vector corresponding to the target's horizontal and vertical coordinates. We assume $\boldsymbol{x}$ follows a two-dimensional Gaussian distribution with mean $\boldsymbol{s}$ and covariance matrix $\frac{1}{J}\mathbf{I}$, where J is a scalar that represents memory precision.

Like in Chapter 1, we adopt the Variable Precision model (van den Berg et al., 2012), assuming that precision J is itself a random variable that follows a gamma distribution with mean $\bar{J}$ and scale parameter τ. Our extension of the model allows the priority-specific $\bar{J}$ to vary; τ is fixed across conditions. We denote the mean total amount of available resource, the sum of the priority-specific precisions, as $\bar{J}_{\text{total}}$.

2.3.2 Resource allocation strategies

In this experiment, we test three models: the Proportional, Flexible, and Minimizing Error model. The models differ in how resource is allocated amongst the different items. We denote the proportion allocated to the high, medium, and low-priority item as p_{high}, p_{med}, and p_{low}, respectively. The mean amount of resource allocated to each item is some proportion of $\bar{J}_{\text{total}}$. For example, the average amount allocated to the high-priority item is $p_{\text{high}}\bar{J}_{\text{total}}$.

2.3.2.1 Proportional model

The Proportional model makes a very strong, unsubstantiated-but-extremely-intuitive hypothesis: people allocate their resource proportional to the probe probabilities. In other words, $p_{\text{high}} = 0.6$, $p_{\text{med}} = 0.3$, and $p_{\text{low}} = 0.1$. Its two free parameters are total resources $\bar{J}_{\text{total}}$ and scale parameter τ.

2.3.2.2 Flexible model

The Flexible model is the complement to the Proportional model: able to describe a wider range of human behaviors, but providing no hypothesis. In this model, the proportions allocated to each priority condition are fitted as free parameters. Thus, this model makes no hypothesis about how observers are allocated resource, only serves to describe what they do. Its four free parameters are then $\bar{J}_{\text{total}}$, τ, p_{high}, and p_{med}.

2.3.2.3 Minimizing error model

The Minimizing Error model is a normative model in which the observer allocates resources in order to minimize expected behavioral cost across the experiment (Bays, 2014; Harris & Wolpert, 1998; Kahneman & Tversky, 1979; Sims, 2015). We assume that the cost on a single trial is related to the magnitude of the estimation error, ϵ, on that trial through a power law:

$$C_{\text{estimation}}(\epsilon) = \epsilon^{\gamma},$$

where $\gamma > 0$. We chose to use a power rather than committing to a type of error function (e.g., absolute or squared error, which is $\gamma = 1$ or 2, respectively). The exponent on the error serves as a "sensitivity to error" parameter: an observer with a large exponent will experience a large error as much more costly than an observer with a lower exponent, and will adjust their strategy accordingly to avoid those errors.

Suppose that on a given trial, the observer has allocated mean resource $\bar{J}$ to the probed stimulus. The expected cost on that trial is then an average over the errors ϵ that could occur on that trial:

$$\begin{aligned} \bar{C}_{\text{estimation}}(\bar{J}) &\equiv \mathbb{E}(C_{\text{estimation}}|\bar{J}) \\ &= \int \epsilon^{\gamma} p(\epsilon|\bar{J}) d\epsilon \\ &= \int \epsilon^{\gamma} \int p(\epsilon|J) p(J|\bar{J}) dJ d\epsilon. \end{aligned} \tag{2.1}$$

Note that we had to additionally marginalize over J, because it is unknown. The analytical derivation of this value can be found in the Supplementary.

So far, we have considered a trial with a given $\bar{J}$. Now, we ask how, for a given $\bar{J}_{\text{total}}$, τ, and γ, should the observer set p_{high}, p_{med}, and p_{low} to minimize the expected cost across the entire experiment. We refer to this expected cost as the "overall expected cost" (OEC); it is equal to

$$\text{OEC}(p_{\text{high}}, p_{\text{med}}, p_{\text{low}}) = 0.6\bar{C}(p_{\text{high}} \cdot \bar{J}_{\text{total}}) + 0.3\bar{C}(p_{\text{med}} \cdot \bar{J}_{\text{total}}) + 0.1\bar{C}(p_{\text{low}} \cdot \bar{J}_{\text{total}}).$$

We denote the resulting cost-minimizing proportions by p^*_{high}, p^*_{med}, and p^*_{low}. The value of probabilities vary as a function of $\bar{J}_{\text{total}}$, τ, and γ.

We assume the observer calculates and uses p^*_{high}, p^*_{med}, and p^*_{low}. We find the values of p^*_{high}, p^*_{med}, and p^*_{low} with fmincon in MATLAB's Optimization Toolbox (MathWorks). We begin the optimization from ten different starting points, to lower the probability of finding a local minimum, and choose the proportions corresponding to the lowest OEC. Note that this optimization is different from the optimization completed to estimate the ML parameters (explained below): the former is necessary to calculate the log-likelihood of a single parameter combination, and is thus completed thousands of times within one ML parameter estimation.

In this Minimizing Error model, resource allocation differs substantially from the Proportional model (De Silva & Ma, 2017). An observer with limited resource should allocate their resource more equally than proportional. Such a strategy would lower the probability of very large errors for low-priority targets, at a small expense of the high-priority targets (Figure 2.2).

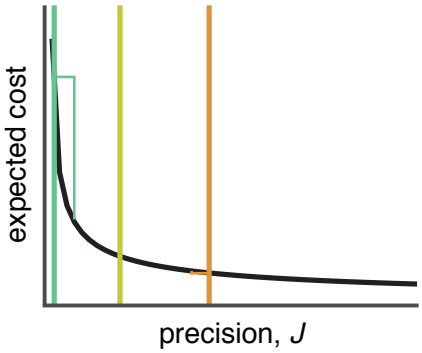

Figure 2.2 Model didactics: Minimizing Error model. The expected cost as a function of precision, J, for the Minimizing Error model. A hypothetical resource allocation choice is illustrated through by the vertical colored lines, indicating the precision associated with the low- (green), medium- (yellow), and high- (orange) items. The overall expected cost (OEC) is the sum of the expected cost for each priority (the value of the cost function at the vertical lines) weighted by their probe probability. The green step illustrates the behavioral benefit for allocating more resource to the low-priority item from the high-priority item; the orange step illustrates the corresponding behavioral detriment to the high-priority item. For this model, the OEC would decrease by allocating slightly more resource to the low-priority item. Thus, optimal allocation favors more equal allocation.

2.3.3 Model prediction

We assume the observer's response, $\hat{\boldsymbol{s}}$, is exactly $\boldsymbol{x}$. This means that the magnitude of the estimation error, $\epsilon \equiv ||\boldsymbol{x} - \boldsymbol{s}||$, follows a Rayleigh distribution with parameter $\frac{1}{\sqrt{J}}$.

The probability of a response $\hat{\boldsymbol{s}}$ given a stimulus $\boldsymbol{s}$ is the following:

$$\begin{aligned} p(\hat{\boldsymbol{s}}|\boldsymbol{s}) &= \iint p(\hat{\boldsymbol{s}} \mid \boldsymbol{x}) p(\boldsymbol{x} \mid \boldsymbol{s}, J) d\boldsymbol{x}\, p(J \mid \bar{J}, \tau) dJ \\ &= \iint \delta(\hat{\boldsymbol{s}} - \boldsymbol{x}) \mathcal{N}\left(\boldsymbol{x}; \boldsymbol{s}, \frac{1}{J}\right) d\boldsymbol{x}\, p(J \mid \bar{J}, \tau) dJ \\ &= \int \mathcal{N}\left(\hat{\boldsymbol{s}}; \boldsymbol{s}, \frac{1}{J}\right) p(J \mid \bar{J}, \tau) dJ \\ &= \int \text{Rayleigh}\left(\epsilon; \frac{1}{\sqrt{J}}\right) p(J \mid \bar{J}, \tau) dJ \end{aligned}$$

We compute this value by numerically integrating over J with 500 equally spaced bins.

2.3.4 Parameter estimation

We use Bayesian Adaptive Direct Search (BADS; Acerbi & Ma, 2017) to estimate, for each participant and model, the parameter combination $\boldsymbol{\theta}$ that maximizes the likelihood of the data given the model. We use 50 different starting positions, using latin hypercube sampling, to minimize the probability of finding a local minimum. We took the maximum of all the runs as our estimate of the maximum-likelihood, and the corresponding parameter combination as our ML parameter estimates.

2.3.5 Parameter and model recovery

To validate the data-generating and model-fitting code, we performed parameter and model recovery. We simulated data from each model then fit each model to the simulated data. Successful parameter recovery occurs when the estimated parameters for the model that generated the data are equivalent or close to the true parameters. Parameter recovery is necessary for the interpretability of the parameter estimates. Successful model recovery occurs when the model which generated the data also fits the data better than any other model. Model recovery is necessary to ensure the models are distinguishable in a psychologically plausible model space. The results of our parameter and model recovery suggest no problems with interpreting parameters of our models or the model comparison.

2.3.6 Model comparison

We compared models using the corrected Akaike Information Criterion (AICc; Hurvich & Tsai, 1987) and the Bayesian Information Criterion (BIC; Schwarz, 1978). Both AICc and BIC penalize models with more parameters, but BIC is more conservative.

2.4 Results

2.4.1 Behavioral Results

Stereotypical eye movements involve one lower accuracy saccade that brings the eye position "into the area of" the target (which we call the primary saccade) and a subsequent saccade that is like an adjustment to the primary saccade (wich we call the final sac-

cade). For every trial, we computed error as the Euclidean distance, in degrees of visual angle, between the true and estimated target location. We conducted three repeated-measures ANOVAs with the priority condition as the within-subject variable and primary saccade error, final saccade error, and reaction time as dependent variables. In line with our hypothesis, primary and final error decreased monotonically with increasing priority ($F(2, 26) = 13.01, p < .001$ and $F(1.18, 15.37) = 10.95, p = .003$, respectively), reflecting the intuition that people allocate more resource to a more behaviorally relevant target (Fig. 2.3). Reaction time decreased slightly with increasing priority ($F(1.3, 17.1) = 3.45, p = .08$). This indicates that the main effect of results is not due to a speed-accuracy trade-off. Because priority effects did not differ between primary and final saccade position, we subsequently only analyze the final saccade data.

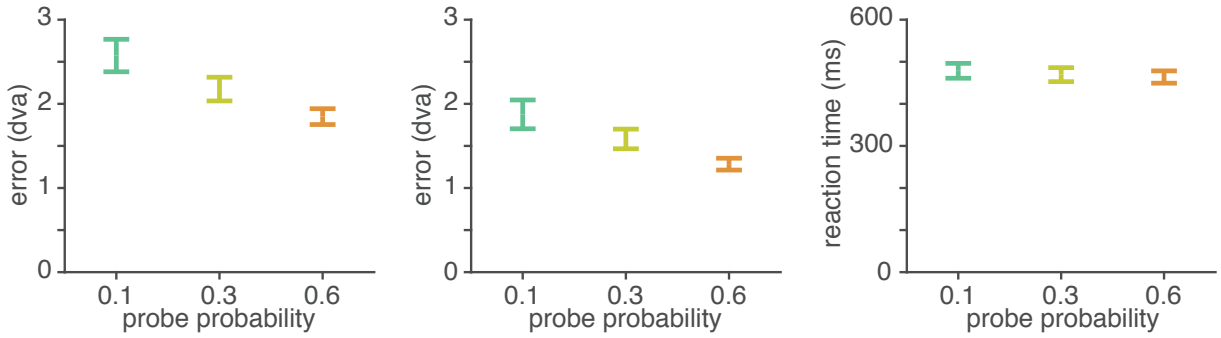

Figure 2.3 Behavioral results. Primary (*left*) and final (*middle*) saccade error ($M \pm SEM$ across participants) decreases as a function of increasing priority. This is effect is not due to a speed-accuracy trade off, since reaction time (*right*) does not increase with priority. green: 0.1, yellow: 0.3, orange: 0.6.

2.4.2 Modeling results

The Proportional model provided a poor fit to the data (left plot of Fig. 2.4A), suggesting that people do not allocate resource in proportion to probe probability. The Flexible

model, on the other hand, fit the data well (middle plot of Fig. 2.4A) and formal model comparison showed that it outperformed the Proportional model by a median AICc of 63 (bootstrapped 95% CI: [37, 107] and a median BIC of 54 [29, 99]; Fig. 2.4B). The proportions allocated to the high-, medium-, and low-priority targets were estimated as 0.49 ± 0.04 ($M \pm SEM$), 0.28 ± 0.02, and 0.23 ± 0.03, respectively (Fig. 2.5), suggesting that the brain underallocates resource to high-priority targets and overallocates resource to low-priority targets, relative to the experimental probe probabilities. While providing an excellent quantitative explanation of the data, it does not provide any explanation for why participants allocate their resource so.

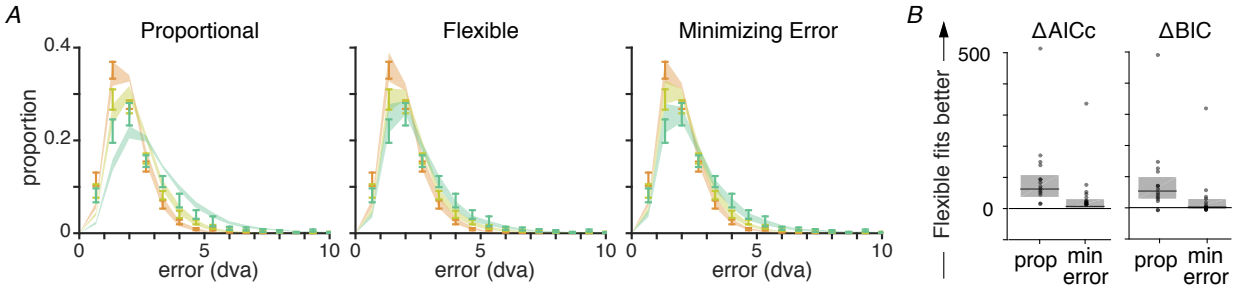

Figure 2.4 Modeling results. *A*. Model predictions. $M \pm SEM$ error distributions for data (error bars) and model predictions (shaded region) for the Proportional, Flexible, and Minimizing Error models ($N = 14$). *B*. Model comparison results. black line: median, grey box: 95% bootstrapped median CI. The Flexible model fits significantly better than the Proportional (prop) model, but not significantly better than the Minimizing Error (min error) model.

The Minimizing Error model (right plot of Fig. 2.4A) fits better than the Proportional model (median AICc [bootstrapped 95% CI]: 49 [21, 99], BIC: 44 [17, 94]) and comparably to the Flexible model (AICc: -7 [-30, -1], BIC: -3 [-26, 3], Fig. 2.4B). Additionally, the model estimated an allocation of resource similar to the Flexible model ($0.46 \pm 0.02, 0.32 \pm 0.01$, and 0.22 ± 0.02 for high-, medium-, and low-priority targets, respectively). This suggests that the under- and over-allocation of resources relative to probe probabilities may be rational, stemming from an attempt to minimize error across the experiment.

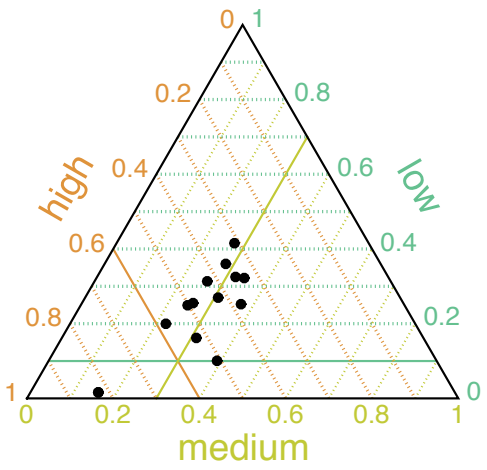

Figure 2.5 Proportion allocated to each priority condition as estimated from the Flexible model. Each black dot corresponds to one participant. Thicker lines indicate the 0.6, 0.3, and 0.1 allocation to high, medium, and low, respectively. The intersection of these lines is the prediction for the Proportional model. Observers are underallocating to high priority and overallocating to low, relative to the actual probe probabilities.

2.5 Discussion

In this study, we investigated the effect of priority on our working memory representations. Specifically, we asked if people could allocate resources across more than two priority levels and, if so, what strategy they were using to decide how much to allocate to each item.

We found that people indeed seemed to be allocating VWM resource according to priority, finding that error decreased significantly and monotonically with increasing priority. This seems to support a truly graded, flexible representation and ability to flexibly allocate resource. While this seems to be strong evidence against a focus or spotlight of attention (Posner et al., 1980; Treisman & Sato, 1990; Oberauer, 2002) and slots+averaging models (Zhang & Luck, 2008), it is possible data under this theory could provide graded representations. For example, a probabilistic spotlight of attention or allocation of slots

may, over trials, results in a monotonic relationship between priority and error. However, our model makes predictions on every trial, and it seems unlikely that a flexible resource model would fit so well if precision was actually two-tiered. An additional model comparison can quantitatively test this.

Through computational modeling, we found that people underallocated resource to the high-priority item and overallocated to the low-priority item, relative to probe probabilities. We were able to describe this in terms of a normative model that assumed people allocated resource in order to minimize behavioral cost, which was defined as estimation error to a power. This model is similar to models proposed by others (Bays, 2014; Sims, 2015; van den Berg & Ma, 2018) and explains the data better than a model that assumes proportional allocation (Klyszejko et al., 2014; Dube et al., 2017), and comparably with the Flexible model, a highly descriptive but not normative model of how people could be allocating resource.

Our results identify a single model of how the resource that supports VWM is allocated, capturing the variability in WM ability across individuals (e.g. Engle, Kane, & Tuholski, 1998; Salthouse, Babcock, & Shaw, 1991) and items (e.g. Adam, Robison, & Vogel, 2018; Reinhart et al., 2012). There are, however, seemingly inconsistent results suggesting that inter-participant variability in WM performance is due to differences in control processes, such as inhibition of irrelevant distractors (Vogel, McCollough, & Machizawa, 2005), may be explained in terms with the "sensitivity to error" parameter, which is the exponent on error when calculating behavioral cost. Some participants prefer making a few large errors in order to maximize the number of extremely precise memory guided saccades, while others prefer avoiding large errors at the expense of those precise saccades.

How would precision be allocated if we changed the cost function of the task? Optimal-

ity is task-dependent, and thus we cannot say people allocate resources optimally without testing different contexts. For example, remembering only a subset of items, relative to all items, may improve performance on one task but detriment it on another (Bengson & Luck, 2016; Atkinson, Baddeley, & Allen, 2018). Would people adjust their allocation strategy according to task demands, or still minimize the behavioral cost of estimation errors?

In the previous and the current chapters, we have demonstrated how people use two pieces of information, uncertainty and priority, in working memory. Specifically, we showed that priority is used during the encoding phase to allocate resource in order to minimize loss and that uncertainty is used during the decision phase to maximize performance. How does priority affect uncertainty? In theory, priority affects precision, which should affect uncertainty. In the next chapter, we address these questions.

2.6 Supplementary

2.6.1 Derivation of $\bar{C}_{\textbf{estimation}}(\bar{J})$

In this section, we derive the analytical expression for calculating the expected estimation cost, for a given $\bar{J}$ on a given trial. Beginning from Eq. 2.1 in the main text.

$$\begin{aligned}
\bar{C}_{\text{estimation}}(\bar{J}) &= \int \epsilon^{\gamma} \int p(\epsilon|J)p(J|\bar{J})dJd\epsilon \\
&= \int \epsilon^{\gamma} \int \epsilon J e^{-\frac{\epsilon^2 J}{2}} \cdot \frac{1}{\Gamma(k)\tau^k} J^{k-1} e^{-\frac{J}{\tau}} dJd\epsilon \\
&= \frac{1}{\Gamma(k)\tau^k} \int \left(\int \epsilon^{\gamma+1} e^{-\frac{\epsilon^2 J}{2}} d\epsilon \right) J^k e^{-\frac{J}{\tau}} dJ \\
&= \frac{\Gamma\left(\frac{\gamma+2}{2}\right)}{2\Gamma(k)\tau^k} \int \left(\frac{2}{J}\right)^{\frac{\gamma+2}{2}} J^k e^{-\frac{J}{\tau}} dJ \\
&= \frac{\Gamma\left(\frac{\gamma+2}{2}\right) 2^{\frac{\gamma+2}{2}}}{2\Gamma(k)\tau^k} \int J^{k-\frac{\gamma}{2}-1} e^{-\frac{J}{\tau}} \\
&= \frac{\Gamma\left(\frac{\gamma+2}{2}\right) 2^{\frac{\gamma+2}{2}}}{2\Gamma(k)\tau^k} \Gamma\left(k-\frac{\gamma}{2}\right) \tau^{k-\frac{\gamma}{2}} \\
&= \frac{\Gamma\left(\frac{\gamma+2}{2}\right) \Gamma\left(k-\frac{\gamma}{2}\right)}{\Gamma(k)} \left(\frac{2}{\tau}\right)^{\frac{\gamma}{2}},
\end{aligned}$$

where $k \equiv \frac{\bar{J}}{\tau}$ and we assumed that $\gamma < 2k$.

2.6.2 Maximum-likelihood parameter estimates

I report summary statistics for only the models that provided reasonable qualitative fits to the data.

	$\bar{J}_{\text{total}}$	τ	p_{high}	p_{med}
mean	3.91	0.77	0.49	0.28
SEM	0.75	0.23	0.04	0.02

Table 2.1 Flexible model parameters. Mean and SEM across participants for all parameters in the Flexible model.

	$\bar{J}_{\text{total}}$	τ	γ
mean	3.97	0.97	0.86
SEM	0.7780	0.2401	0.2303

Table 2.2 Minimizing Error model parameters. Mean and SEM across participants for all parameters in the Minimizing Error model.

3 Consistent strategy use across different behavioral contexts

Consistency is the last refuge of the unimaginative.

Oscar Wilde

3.1 Introduction

In Chapter 1, we found that people maintained and used an accurate representation of uncertainty when making change detection decisions. This provided evidence of probabilistic computation in working memory. In Chapter 2, we found that people used priority information to strategically allocate resource, and they did so in a way that was consistent with a loss-minimizing strategy. In this chapter, we investigate uncertainty and priority together, asking if we can replicate and generalize the results of previous chapters in new contexts.

In this task, we ask three main questions. First, would priority affect uncertainty representations? In Chapter 1, we manipulated the mean precision of different ellipses through the amount of bottom-up information available during stimulus presentation, namely the ellipse eccentricity. We found that people's uncertainty reflected the precision with which each item was remembered. In Chapter 2, we showed that we could also manipulate the main precision of items through a top-down manipulation, namely priority. Do people's representations of uncertainty truly reflect the precision with which they are encoded, or are they sensitive to different external manipulations of precision?

Second, would people use uncertainty optimally when making working memory decisions? In Chapter 1, people's behavior was well fit by a model that assumed they combined information optimally. However, a model that did not assume optimality fit the data comparably well. We wanted to test, in an entirely different task, if people would perform in a way consistent with the optimal strategy.

Finally, would people change their encoding strategy based on task demands? In Chap-

ter 2, participants allocated resource to minimize a function of saccade error. If we incentivized a different strategy, would people adjust their strategy or continue to minimize error?

To investigate these questions, we conducted the same delayed-estimation task as in Ch. 2, with an additional post-estimation wager. The goal of the post-decision wager was to capture the true target location with a circle that was centered around the participants' saccade landing. The participant could adjust the radius of the circle. Points were awarded if the true target was within the circle, and smaller circles were awarded more points. The addition of this wager was crucial to test our hypothesis.

For the first question, we predicted that priority would affect memory precision, which would in turn affect the uncertainty people have in the accuracy of their memory-guided saccade. More concereteley, we predicted people would have, on average, smaller circle sizes for higher priority items. However, we believe memory precision fluctuates on an item-specific basis, not only through priority. Thus, we also predict circle size would vary according to item-specific fluctuations in memory precision, as measured through saccade error.

Like confidence reports (e.g. van den Berg, Yoo, & Ma, 2017; Rademaker et al., 2012), the reported circle size in the wager should be related to memory precision. Unlike confidence reports, we can test optimal use of uncertainty with the reported circle size. This is because there is a performance-maximizing circle size an observer should set based on the cost function and their uncertainty. Note that the optimal use of uncertainty occurs during the wager decision stage and is based on the current precision.

This is in contrast to the optimal allocation of resource (our third question), which oc-

curs during encoding. In this task, participants are instructed to maximize the amount of points they earn. Interestingly, the strategy to maximize points results in different resource allocation than the strategy to minimize estimation error, which allows us to quantitatively test whether people are changing their strategies based on a different experimental contexts or if they are continuing to minimize error.

3.2 Experimental methods

3.2.1 Participants

Eleven people (5 males, age=28.6, SD=3.03) participated in this experiment. Everyone had normal or corrected-to-normal vision and no history of neurological disorders. Participants were naive to the study hypotheses and were paid $10/hour. Participants completed one or two one-hour sessions. We obtained informed, written consent from all participants. The study was in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of New York University.

3.2.2 Apparatus

Participants were placed 56 cm from the monitor (19 inches, 60 Hz), with their heads in a chinrest. Eye movements were calibrated using the 9-point calibration and recorded at a frequency of 1000 Hz (Eyelink 1000, SR Research). Target stimuli were programmed in MATLAB (MathWorks) using the MGL toolbox (Gardner Lab, Stanford) and were displayed against a uniform grey background.

Participants made behavioral responses using a space bar with their left hand and a circular knob (PowerMate, Griffin Technology) with their right hand. For eye-tracking, we applied an online drift correction when the recorded location of center of fixation exceeded 1 degree of visual angle (dva) from the center of the fixation cross. This was because this experiment provided live visual feedback of the participants' current fixation and large discrepancies were uncomfortable for the participant and resulted in imprecise data.[1]

3.2.3 Trial procedure

The trial sequence in this experiment was identical to the experiment in Chapter 2 until the response cue. Each trial (Fig. 3.1) began with a 300 ms increase in the size of the fixation symbol, an encircled fixation cross. This was followed by a 400 ms endogenous precue, consisting of three colored wedges presented within the fixation symbol, each of which angularly filled one quadrant. The radial sizes and colors of the wedges corresponded to probe probabilities of 0.6, 0.3, and 0.1, respectively. (While visualized as being orange, yellow, and green wedges, the actual colors were pink, yellow, and blue). The quadrant with a probe probability of 0.0 did not have a wedge.

The precue was presented for 400 ms; this was followed by a 700 ms interstimulus interval, then by the targets, presented for 100 ms. The targets were four dots, each in separate visual quadrants. The dots were presented at approximately 10 degrees of visual angle from fixation, with random jitter of 1 degree of visual angle to each location. The location of the targets in polar coordinates were pseudo-randomly sampled from every 10 degrees, avoiding cardinal axes.

[1] This was not a problem in the experiment in Chapter 2 because the corrective saccade provided a measure of drift, which we used to correct offline.

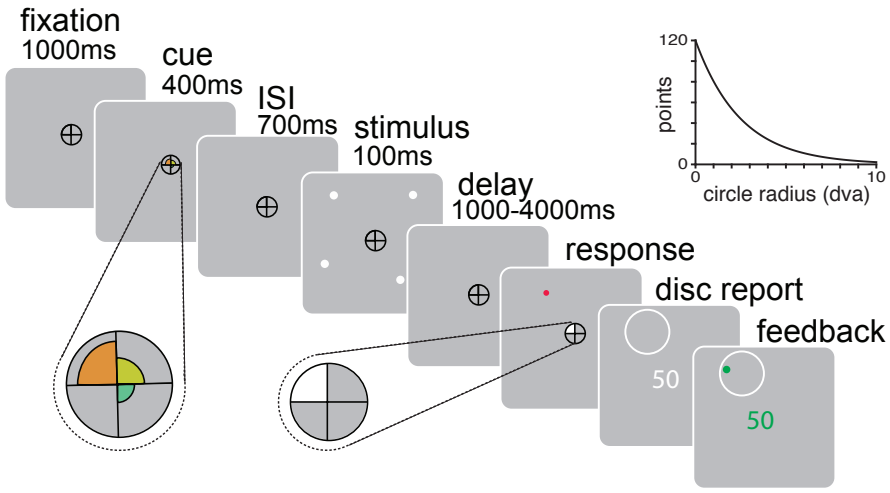

Figure 3.1 Trial sequence. Participants saw a precue indicating the probe likelihoods of each of the items, maintained the location of them over a delay period, made a memory-guided saccade to the remembered location of a probed dot, then completed a post-decision wager. In the wager, participants adjusted the radius of the circle, with the goal of enclosing the true target within the circle. They were only awarded points if the true target location was within the circle, with the amount of points decreasing exponentially with circle radius. Points awarded as a function of circle radius shown in top right.

This was followed by a variable delay, chosen with equal probability from the range between 1000 and 4000 ms in 500 ms increments. A response cue appeared afterward, which was a white wedge that filled an entire quadrant of the fixation symbol. Simultaneously with the response cue, a red dot appeared at the location of the participants' fixation as measured online by the eye tracker (Graf, Warren, & Maloney, 2005). Because of eye-tracker noise, the red dot occasionally appeared in a slightly different location than where the participant was fixating. In these cases, participants were instructed to adjust their gaze such that the red dot was at the remembered location, and press the space bar to indicate that this was their intended saccade endpoint. After completing this response, participants performed a post-estimation wager. A circle appeared, centered at the saccade endpoint. Participants received points based on the size of the circle, such that a smaller circle corresponded to more points. Participants were only rewarded points if the true target was within the circle. The amount of potential points awarded were displayed as participants changed the radius of the circle. The number of points awarded was $120e^{-0.4r}$, in which r was the radius of the circle in dva (this is illustrated in the top right of Fig. 3.1). The circle size report served as a measure of memory uncertainty because participants were incentivized to make smaller circles when their memory was more certain, in order to obtain more points. Participants were incentivized to maximize points with additional monetary reward.

Participants completed between 8 and 19 forty-trial runs over one or two one-hour sessions.

3.2.4 Data processing

Processing and manual scoring of eye movement data was performed in an in-house MATLAB function-graphing toolbox (iEye). Eye position was extracted from iEye. We excluded trials in which a) participants were not fixating in the middle of the screen during stimulus presentation, b) saccades were initiated before 100 ms or after 1200 ms after the response cue onset, c) pupil data during the response period were missing, d) participants made a saccade to the wrong quadrant, ignoring the response cue, or e) primary saccade error was over 10 dva. After data exclusion, there was a total of between 320 and 760 trials were completed across participants.

3.3 Modeling methods

The model prediction, fitting, and comparison methods are closely aligned with those in Chapter 2. We reiterate our methods below, but more explanation and justification can be found in the Chapter 2 Section 2.3.

3.3.1 Encoding and decision stage

In this experiment, we obtained two measures of behavior on every trial: the memory estimation and the circle wager. We model target location $\mathbf{s}$ as a two-dimensional vector corresponding to the target's horizontal and vertical coordinates. We denote the observer's estimate (saccade endpoint) of $\mathbf{s}$ by $\mathbf{x}$. We assume $\boldsymbol{x}$ follows a two-dimensional Gaussian distribution with mean $\mathbf{s}$ and covariance matrix $\frac{1}{J}\mathbf{I}$, where J is a scalar. We assume that the observer's estimation, $\hat{\boldsymbol{s}}$ is exactly $\boldsymbol{x}$. We conveniently ignore motor and response noise

(and well known saccade biases), recognizing that the variability in the VP model may account for these.

We assume that on every trial, the observer chooses a circle radius r noisily around the value that maximizes the expected utility (EU) of that trial. The EU is calculated as the product between the utility of setting a circle with radius r and the probability that the true stimulus lies within the circle bounds (i.e., a hit). We define utility as the number of points awarded for circle radius r, raised to an exponent α that accounts for risk preferences, $120e^{-r\alpha}$. An $\alpha > 1$ corresponds to risk-seeking behavior (corresponding to smaller circles than the optimal observer would set), while an $\alpha < 1$ corresponds to risk-averse behavior (corresponding to larger circles than the optimal observer would set).

The probability of a hit is equivalent to the bounded integral of the posterior $p(s|x)$ over the region described by the circle. For a two-dimensional Gaussian distribution, this is equivalent to a cumulative Rayleigh distribution evaluated at r:

$$\begin{aligned} p_{\text{hit}}(r, J) &\equiv p(\epsilon \leq r|J) \\ &= \left(1 - e^{-\frac{r^2 J}{2}}\right). \end{aligned}$$

We assume that the observer's noisily choose the optimal circle radius, following a softmax rule. The probability of choosing r is

$$\begin{aligned} p(r|J) &\propto \exp\left(\beta \text{EU}(r, J)\right) \\ &= \exp\left(\beta \cdot \text{utility}(r) \cdot p_{\text{hit}}(r, J)\right) \\ &= \exp\left(\beta \cdot 120e^{-r\alpha}\left(1 - e^{-\frac{r^2 J}{2}}\right)\right). \end{aligned}$$

The observer is thus most likely to choose the maximum EU, with some noise determined by β, the inverse temperature parameter: a lower β corresponds to higher decision noise. In this formulation, we are assuming that the observer knows the point function. This may be a reasonable assumption because participants completed training on the task, and the potential points awarded were always presented when participants were making the post-estimation wager decision.

3.3.2 Resource allocation strategies

As defined in Chapter 2 Section 2.3.2, we test the Proportional, Flexible, and Minimizing Error model. We additionally test the optimal model for this task, which is to maximize the amount of points earned on the wager.

3.3.2.1 Proportional model

In the Proportional model, observers allocate resources equivalently to the experimental probe probabilities, i.e., $p_{\text{high}} = 0.6, p_{\text{med}} = 0.3, p_{\text{low}} = 0.1$. Its four free parameters are total resources $\bar{J}_{\text{total}}$, scale parameter τ, risk preference α, and inverse noise temperature β.

3.3.2.2 Flexible model

In the Flexible model, the proportions allocated to each priority condition are fit as free parameters. Its six free parameters are then $\bar{J}_{\text{total}}$, τ, α, β, p_{high}, p_{med}.

3.3.2.3 Minimizing Error model

In the Minimizing Error model, observers allocate resource in order to minimize expected behavioral loss across the experiment exactly as described in Chapter 2. While perhaps optimal in the earlier study, this strategy is myopic for the current experiment: the observer does not take into account the subsequent decision they must make, but first maximizes performance in terms of estimation error, then maximizes EU. Its five free parameters are $\bar{J}_{\text{total}}$, τ, γ, α, and β. The optimal resource allocations $p^*_{\text{high}}, p^*_{\text{med}}, p^*_{\text{low}}$ depend only on parameters $\bar{J}_{\text{total}}$, τ, and γ.

3.3.2.4 Maximizing Points model

While observers in all models maximize the EU on every trial for a given J, the Maximizing Points model observer additionally allocates resources in order to maximize the expected utility across the entire experiment. We define the cost of a single trial as the negative EU on that trial:

$$C_{\text{wager}}(r|J) = -\left(1 - e^{-\frac{r^2 J}{2}}\right) \cdot 120e^{-r\alpha}.$$

The expected cost on that trial, for a given J, is an average of the costs for all possible radii r reported on that trial multiplied by its probability. However, J itself is a random variable, drawn from a distribution determined by a priority-specific $\bar{J}$. Thus, we must

also marginalize over J to calculate the expected cost of a trial in each priority condition:

$$\begin{aligned}\bar{C}_{\text{wager}}(\bar{J}) &\equiv \mathbb{E}(C_{\text{wager}}|r, J) \\ &= \int C_{\text{wager}}(r|\bar{J})p(r|J)dr \\ &= \iint -\text{EU}(r, J)p(r|J)p\left(J \,\middle|\, \bar{J}, \tau\right) drdJ\end{aligned}$$

We numerically integrated over r and J to obtain the $\bar{C}_{\text{wager}}$ for a given $\bar{J}$. The overall expected cost (OEC) for this experiment is thus:

$$\text{OEC}(p_{\text{high}}, p_{\text{med}}, p_{\text{low}}) = 0.6\bar{C}(p_{\text{high}} \cdot \bar{J}_{\text{total}}) + 0.3\bar{C}(p_{\text{med}} \cdot \bar{J}_{\text{total}}) + 0.1\bar{C}(p_{\text{low}} \cdot \bar{J}_{\text{total}}). \quad (3.1)$$

In the Maximizing Points model, the cost-minimizing proportions $p^*_{\text{high}}, p^*_{\text{med}}$, and p^*_{low} are a function of all parameters $\bar{J}_{\text{total}}$, τ, α, and β. We obtain these values through the optimization methods described in Chapter 2.

3.3.2.5 Comparison of different models

In order to find the optimal resource allocation, the Minimizing Error and Maximizing Points observers must minimize the overall expected cost. For both models, this value is the sum of the expected error for each priority, weighted by their probe probability (Equation 3.1). The weight reflects the true frequency of times each priority item is probed throughout the experiment. Finding the minimum OEC becomes somewhat of a balancing game, where the cost for one item will be lowered at the expense of the other, and the cost of high-priority items contributes 2 and 6 times more to the OEC than the medium-

and low-priority items do, respectively. Figure 3.2A shows how each model minimizes the overall expected cost, and how their predictions differ based on the shape of this function. For example, for the Minimizing Error model, the OEC would decrease substantially by allocating slightly more resource to the low-priority item from the high-priority item. Thus, optimal allocation according to the Minimizing Error model may favor more equal allocation than proportional. The Maximizing Points model would not benefit from this same trade off. You can see in the diagram that cost of the low-priority item would decrease more than the high-priority item would increase, but would not lead to a lower OEC because the high-priority item is weighted so much more. Optimal allocation according to the Maximizing Points model thus results in allocating most of the resource to the high priority item at the expense of the low priority item, compared to proportional. These trends are true for low total precision $\bar{J}_{\text{total}}$; the optimal allocation moves toward proportional allocation with higher $\bar{J}_{\text{total}}$ (Fig. 3.2B).

3.3.3 Model prediction

We predict the probability of the memory estimation $\hat{\boldsymbol{s}}$ and wager circle size r given the stimuli. We assume these two data points are conditionally independent.

$$\begin{aligned} p(\hat{\boldsymbol{s}}, r \mid \boldsymbol{s}) &= \iint p(\hat{\boldsymbol{s}} \mid \boldsymbol{x}) p(\boldsymbol{x} \mid \boldsymbol{s}, J) d\boldsymbol{x}\; p(r \mid J) p(J \mid \bar{J}, \tau) dJ \\ &= \int p(\hat{\boldsymbol{s}} \mid \boldsymbol{s}, J) p(r \mid J) p(J \mid \bar{J}, \tau) dJ \\ &= \int \text{Rayleigh}\left(\epsilon; \frac{1}{\sqrt{J}}\right) \exp\left(\beta \cdot 120 e^{-r\alpha}\left(1 - e^{-\frac{r^2 J}{2}}\right)\right) p(J \mid \bar{J}, \tau) dJ \end{aligned}$$

We compute this value by numerically integrating over J with 500 equally spaced bins.

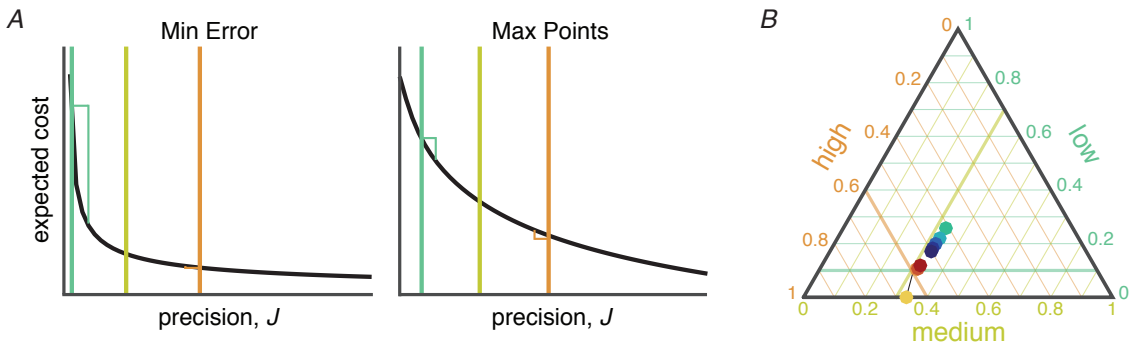

Figure 3.2 Model didactics: comparing optimal resource allocation of Minimizing Error and Maximizing Points models. *A.* The expected cost as a function of precision, J, for the Minimizing Error (*left*) and Maximizing Points model (*right*). A hypothetical resource allocation choice is illustrated through by the vertical colored lines, indicating the precision associated with the low- (green), medium- (yellow), and high- (orange) items. The overall expected cost (OEC) is the sum of the expected cost for each priority (the value of the cost function at the vertical lines) weighted by their probe probability. The green step illustrates the behavioral benefit for allocating more resource to the low-priority item from the high-priority item; the orange step illustrates the corresponding behavioral detriment to the high-priority item. For the Minimizing Error model, the OEC would decrease by allocating slightly more resource to the low-priority item. Thus, optimal allocation according to the Minimizing Error model favors more equal allocation. For the Maximizing Points model, OEC would increase by allocating slightly more resource to low-priority item from the high priority item. Thus, optimal allocation according to the Maximizing Points model favors more extreme differences in proportions allocated. *B.* Optimal resource allocation as a function of increasing total resource, $\bar{J}_{\text{total}}$, for Minimizing Error (green to purple) and Maximizing Points model (yellow to red). Each side of the triangle corresponds to the probe probability of or proportion allocated to each priority condition. With a lower $\bar{J}_{\text{total}}$ (illustrated by the green dot for Minimizing Error model and the yellow dot for Maximizing points model), the models make very different predictions. The Minimizing Error model predicts more equal allocation than proportional to the probe probability, while the Maximizing Points model predicts dropping the low-priority target. As $\bar{J}_{\text{total}}$ increases, both models predict an allocation closer to the experimental probe probabilities.

3.3.4 Parameter estimation

We use Bayesian Adaptive Direct Search (BADS; Acerbi & Ma, 2017) to estimate, for each participant and model, the parameter combination $\boldsymbol{\theta}$ that maximizes the likelihood of the data given the model. We use 50 different starting positions, using latin hypercube sampling, to minimize the probability of finding a local minimum. We took the maximum of all the runs as our estimate of the maximum-likelihood, and the corresponding parameter combination as our ML parameter estimates.

3.3.5 Parameter and model recovery

To validate the data-generating and model-fitting code, we performed parameter and model recovery. The results of our parameter and model recovery suggest no problems with interpreting parameters of our models or the model comparison.

3.3.6 Model comparison

We compared models using the corrected Akaike Information Criterion (AICc; Hurvich & Tsai, 1987) and the Bayesian Information Criterion (BIC; Schwarz, 1978). Both AICc and BIC penalize models with more parameters, but BIC is more conservative.

3.4 Results

3.4.1 Behavioral results

Our predictions for this experiment were the following: a) estimation error decreases with increasing priority, b) circle size decreases with increasing priority, and c) estimation error correlates positively with circle size within each priority level. To test the first two predictions, we conducted a repeated-measures ANOVA with priority condition as the within-subject variable. The ANOVA for circle size violated the assumption of sphericity, so we implemented a Greenhouse-Geisser correction. We conducted Spearman correlations for each priority condition, computing correlations across participants as well as for individual participants. For the aggregated correlation, we removed any participant-specific main effects by standardizing the data ($M = 0, SD = 1$) for each participant before aggregating data for each priority condition.

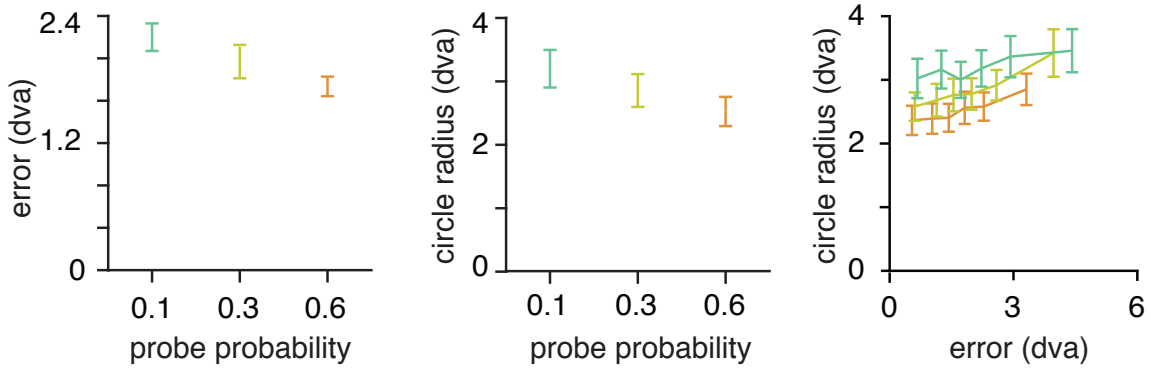

Figure 3.3 Main behavioral results. Error bars show $M \pm SEM$ for memory error (*left*) and circle radius (*middle*) across priorities for 11 participants; both measures decrease with increasing priority. These measures are positively correlated within priority conditions (*right*), suggesting that error and circle size have a common cause, namely fluctuations in precision.

We confirmed all three predictions. First, estimation error decreased monotonically with increasing priority ($F(2, 20) = 12.5, p < 0.001, \eta^2 = 0.55$; left plot of Fig. 3.3), indicating that participants allocated more resource to higher priority targets. Second, the radii of the circle wagers decreased monotonically with increasing priority ($F(1.3, 12.9) = 10.60, p < 0.005, \eta^2 = 0.51$; middle panel of Fig. 3.3), indicating that participants had higher memory certainty in higher priority trials. Third, estimation error and circle size were correlated within each priority level across participants ($\rho_{0.6} = 0.22, p < 0.001; \rho_{0.3} = 0.28, p < 0.001; \rho_{0.1} = 0.18, p < 0.001$; right panel of Fig. 3.3), indicating that people have a single-trial representation of their uncertainty independent of the priority manipulation, as suggested by earlier work (Fougnie et al., 2012; Suchow et al., 2017). Correlations at the individual level resulted in similar correlation values ($M \pm SEM : \rho_{0.6} = 0.22 \pm 0.04, \rho_{0.3} = 0.27 \pm 0.03, \rho_{0.1} = 0.16 \pm 0.04$), though not all correlations were significant.

These correlations, however, could be driven by some other factor, such as stimulus location or trial delay time. For example, in orientation perception, targets with orientations closer to the cardinal axis are perceived more accurately than obliquely-oriented objects (Appelle, 1972; Furmanski & Engel, 2000; Girshick, Landy, & Simoncelli, 2011; Pratte, Park, Rademaker, & Tong, 2017). There could be a similar effect in working memory, and knowledge of this effect could be driving the measured within-priority correlation. An effect of delay on error, and knowledge of this, could also be driving the correlation. To test these hypotheses, we conducted two permutation tests for each participant and priority level.

First, we completed a regression to see if there was a relationship between either distance from cardinal axis (up to 45°) or delay and estimation error. The oblique effect in memory

of locations of objects was inconclusive. For seven of eleven participants, stimulus location did not significantly predict error ($p > 0.05$), but the remaining four participants had greater error when moving farther from the cardinal axes ($M \pm SEM$ regression weights: $1.40 \pm 0.84, p < 0.05$). Delay did not significantly predict error for nine of eleven participants ($p > 0.05$), and predicted an 0.10 and 0.11 dva increase in error for every second increase in delay for the other two participants ($p < 0.01$).

Because there is not a clear relationship between these variable and error, we cannot simply regress out any relationship between the two. However, there may still be some stimulus-dependent relationship that can still be driving the correlation between error and circle size. We decided to conduct a permutation test, which allows us to investigate this question without needing to describe or parameterize the relationship between the variables of interest. Below, I explain the permutation test to test if stimulus location affected the correlation. For each participant and priority condition:

1. Bin error and circle size data according the angular distance of the target from the horizontal axis (10° to 80° in 10° increments. Note that 10°, 170°, 190°, and 350° are all 10° away from horizontal axis)
2. Permute circle size within each bin
3. Combine bins
4. Compute correlation between error and circle size Repeat steps 2 through 4 a thousand times

In step one, we combined data according to their angular distance from horizontal in an effort to increase the number of trials per bin. This grouping assumes that the main

stimulus-dependent noise would be relative to the cardinal axes, not any hemispheric differences. In an ideal scenario, we would then be able to conduct a correlation within each bin. However, there were as few as two trials in one bin, so computing a correlation for each bin was not feasible.

For step two, we performed a special type of permutation called a derangement, in which no element is placed its original location. We conducted a derangement because it is more robust to small sample sizes than a regular permutation. For example, in a regular permutation of two data points, half of the time you would get the original configuration, leading to biased results.

By completing permutations on multiple small bins within each dataset, the recombined, permuted dataset maintains any correlations that are stimulus-location driven, while removing any relationship driven by a knowledge of internal memory fluctuations. Therefore, if the correlation was largely due to the stimulus location, then the correlation of the permuted data would still be positive. If, on the other hand, the correlation was driven by internal fluctuations that were independent of the location of the stimulus, the positive correlation observed in the non-permuted data would be significantly reduced in the permuted data.

I completed a very similar permutation test to test an effect of delay on the correlation, by binning data by participant, priority, and delay time (1 to 4 seconds in 0.5 second increments); deranging the circle sizes within each bin; combining data across delay bins; then computing the correlation between error and circle size, resulting in a correlation for each participant and priority condition. I repeated this process 1000 times, to get a null distribution of correlation coefficients.

To test if the true correlation was significantly higher than the null correlations, we conducted a Wilcoxon signed-rank test between the medians of each null correlation distribution (for each priority and subject) and the respective true correlations. The actual correlations ($M \pm SEM$: 0.29 ± 0.04) were significantly higher than the median of the correlations obtained in the null distribution when permuting based on stimulus location ($M \pm SEM$: -0.007 ± 0.006; Wilcoxon signed-rank test, $z = -4.69, p < 1e-5$) or delay time ($M \pm SEM$: $-0.004 \pm 0.004; z = -4.53, p < 1e-5$), suggesting that the correlation within each priority condition was driven by knowledge of internal fluctuations in the quality of the memory representation above and beyond any location- or delay-dependent variation.

3.4.2 Modeling results

We again tested the Proportional model and Flexible model, jointly fitting the estimation data and the post-estimation wager data. We again found that the Proportional model did not provide a good fit to human data and the Flexible model provided an excellent fit to the data (first two columns of Fig. 3.4). As before, the Flexible model suggests that the brain underallocates resource to high-priority targets and overallocates resource to low-priority targets relative to experimental probe probabilities. The proportion allocated to the high-, medium-, and low-priority targets were estimated as $0.44 \pm 0.02, 0.31 \pm 0.02$, and 0.25 ± 0.02, respectively (Fig. 3.5).

Unlike in the first experiment, optimal performance in this experiment requires maximizing points. This Maximizing Points model has qualitatively different properties from the Minimizing Error model. An observer that maximizes points would receive more points by ignoring the low-priority targets completely in order to remember the high-priority targets

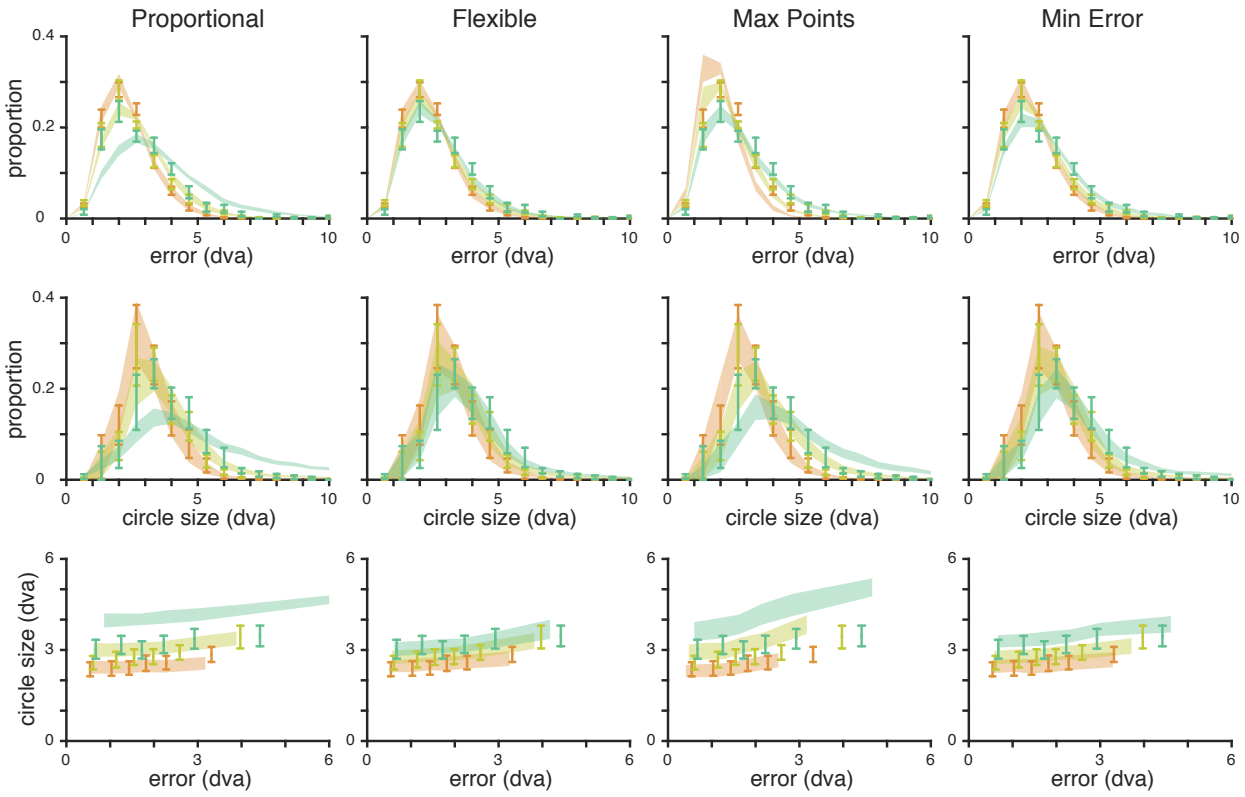

Figure 3.4 Model predictions. Color indicates priority condition — orange: 0.6, yellow: 0.3, green: 0.1. Fits of four models (columns) to error distribution (*top*), circle radius distribution (*middle*), and correlation between the two (*bottom*). $M \pm SEM$ shown for data (error bars) and model predictions (shaded region).

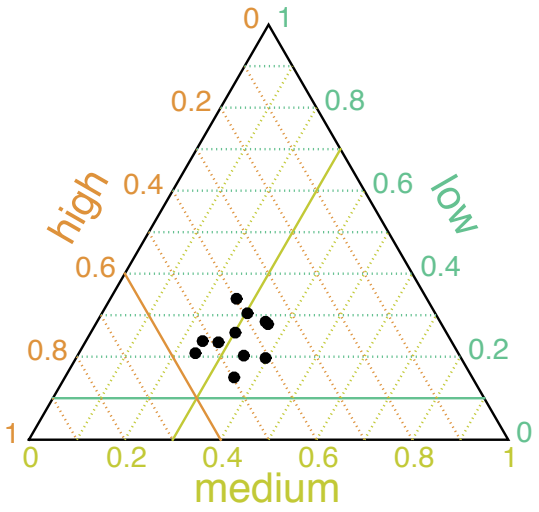

Figure 3.5 Proportion allocated to each priority condition as estimated from the Flexible model. Each black dot represents one participant. Thicker lines indicate the 0.6, 0.3, and 0.1 allocation to high, medium, and low, respectively. The intersection of these lines is the prediction for the Proportional model. Again, observers are underallocating to high priority and overallocating to low, relative to the actual probe probabilities.

better, while an observer that minimizes error would allocate it more evenly across targets. Because these two strategies conflict, we are able to test whether the intrinsically-driven, error-minimizing strategy that people seem to be using in the absence of reward can withstand being put in conflict with an external incentive. The Maximizing Points model fit very poorly, indicating that participants were not allocating resource in order to earn the most points (Proportional model: median ΔAICc: -75 [-109, -26], ΔBIC: -75 [-109, -26]; Flexible model: ΔAICc: -156 [-308, -94], ΔBIC: -148 [-300, -86]; third column Fig. 3.4).

The Minimizing Error model fit the data substantially better than the Proportional model (median ΔAICc: 55 [20, 106], ΔBIC: 50 [17, 102]) and the Maximizing Points model (ΔAICc: 140 [85, 249], ΔBIC: 137 [81, 245]) and about as well as the Flexible model (ΔAICc: -16 [-44, -5], ΔBIC: -12 [-40, 0]; fourth column Fig. 3.4). The Minimizing Er-

ror model fitted the proportions of resource allocated to high-, medium-, and low-priority targets as $0.52 \pm 0.02, 0.32 \pm 0.01$, and 0.16 ± 0.01, respectively, similar to the allocation estimated in the Flexible model. Formal model comparison shown in Figure 3.6.

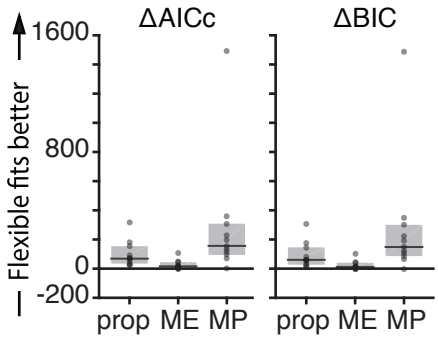

Figure 3.6 Model comparison results. Model comparison difference between the Flexible model and all other models. dots: individual participant model differences. black line: median, grey box: 95% bootstrapped median CI. The Flexible model fits significantly better than the Proportional and Maximizing Points (MP) models, but not significantly better than the Minimizing Error (ME) model.

3.5 Discussion

In this chapter, we study both uncertainty and priority, asking if we could replicate and generalize the results found in the previous two chapters. Specifically, we asked if priority affected uncertainty representations, if people would use uncertainty optimally in another task, and if changing the context in which priority was used would lead to different allocation strategies. Participants completed a four-item memory-guided saccade task with a post-estimation wager.

3.5.1 Summary of results

First, priority affected the precision with which items were remembered, with remarkable consistency to the effects found in Ch. 2. We found that participants maintained a representation of uncertainty that was correlated with error within each priority condition. This correlation could not be explained by stimulus location or delay time, suggesting people's circle size truly represented their item-specific uncertainty. This corroborates that people's confidence typically track with accuracy (e.g., Rademaker et al., 2012; Suchow et al., 2017; van den Berg et al., 2017).

Second, we found that participants' circle size data was well described by an optimal Bayesian model, in which observers combined uncertainty information with the cost function in order to maximize performance on the wager task. There are a couple limitations to the interpretation of this results. First, we did not test any other models for how someone could be making this decision, so it is possible that data could be better described by a suboptimal model. Second, we assume participants learned the cost function, but it is possible that people did not. However, we were able to show in two different tasks (Ch. 1 and here) with two different cost functions and decisions that people maintained and used uncertainty in a way that was consistent with the Bayes-optimal observer.

Third, we found that people continued to be best fit by the Minimizing Error model, despite performance on this task being defined by the points earned on the wager. In fact, the model comparison results were almost identical to that of Chapter 2; the Minimizing Error model fit just as well as the Flexible model and substantially better than the Proportional model. However, the fact that the Flexible model fit well in Ch. 2 does not trivially imply that it would fit the data well here. We made several assumptions when

extending the model to account for circle size data that did not need to apply to human data. First, we assumed that error and uncertainty are generated from the same precision. Second, we assumed people knew their uncertainty. Finally, we assumed that observers set the optimal circle size, given a precision.

Why might people be minimizing error instead of maximizing points? Additionally, in both experiments in this and the previous chapter, we provided feedback on the actual location of the target. This may have provided information for participants to learn a mapping between resource allocation and error.

However, people may be behaving rationally. Perhaps people find minimizing the errors of our memory intrinsically rewarding. Indeed, extrinsic rewards influence the metrics of saccades in humans and monkeys (Chen, Chen, Zhou, & Mustain, 2014; Takikawa, Kawagoe, Itoh, Nakahara, & Hikosaka, 2002). For instance, extrinsic rewards affect both the velocity of saccades as well as neural activity in dopamine-associated reward circuits (Kato et al., 1995), and they modulate neural activity in cortical areas that represent the goals of saccade plans (Platt & Glimcher, 1999). Perhaps the intrinsic reward associated with veridical memory eclipses the extrinsic reward associated with gaining more points, which would explain why people minimized error instead of maximized points in the second experiment. It would be interesting to see if we could have biased behavior toward the Maximizing Points strategy by providing a leader board or giving a monetary bonus based on the performance in the post-estimation wager.

Additionally, minimizing memory error might be computationally easier than maximizing points because it does not require the observer to think and optimize performance two steps ahead. The amount that performance may improve from maximizing points may not be worth the computational and metabolic cost.

3.5.2 Limitations and future work

While the Minimizing Error model provided a good fit to the data, it is possible that people could be doing something much simpler. For example, these data might be explained by a regression toward the mean or probability distortion (Kahneman & Tversky, 1979). Additionally, we tested only power law functions, but more robust loss functions, for example Huber loss (Huber, 1964) could describe data better. Future studies should investigate if resource allocation abides by the Minimizing Error model across a variety of experimental probe probabilities and reward contexts (e.g., monetary reward in Klyszejko et al., 2014).

In this and the previous chapter, we only compared models using AICc and BIC of the maximum-likelihood estimation. All models are nested within the Flexible model, so the Flexible model is guaranteed to fit as well as any other model, in terms of the maximum log-likelihood. AICc and BIC may not sufficiently penalize this model and thus bias results to favor this model. Indeed, I found in model recovery that the Flexible model sometimes was favored in datasets generated from the Minimizing Error model[2]. Perhaps comparing data using marginal likelihoods or cross validation, in which model flexibility is punished more, would provide a better measure of parsimony.

In our experiment, we have a no-priority item which is never probed. Do people still maintain a representation of this? I think so. People use configurational structure when remembering stimuli (e.g. Brady & Tenenbaum, 2013), even though we ignore this aspect

[2] I don't mean to lack transparency for not reporting these results in my dissertation. I simply ran out of time! Model recovery was successful except that the Flexible model fit better for around half of the datasets generated by the Minimizing Error model. However, that doesn't affect the interpretation of the modeling results.

in our computational models. People may use compression strategies that, while facilitating memory, may also produce behavioral biases (e.g. Nassar, Helmers, & Frank, 2018). We could exploit these effects. For example, we could design a task in which participants make a response that depends on all remembered stimuli (e.g., ensemble statistics), and see if the responses are influenced by the no-priority item. With neuroimaging, we could investigate whether there is a representation of the no-priority item during a WM delay. I am interested in pursuing this question in the future.

3.5.3 Conclusions

In this chapter, we linked concepts from previous chapters, asking about how priority and uncertainty interacted in a working memory task. We found remarkable consistency between the previous and current results, finding that priority affects memory precision and that people maintain and use uncertainty in a way consistent with the optimal strategy. We additionally replicated that people allocate resource consistent with a loss-minimizing strategy, although that was not the optimal strategy in this task. In the next chapter, we attempt to begin linking some of these computational concepts in the brain, by asking how priority is represented in the brain. during a working memory delay.

3.6 Supplementary

3.6.1 Maximum-likelihood parameter estimates

I report summary statistics for only the models that provided reasonable qualitative fits to the data.

	$\bar{J}_{\text{total}}$	τ	α	β	p_{high}	p_{med}
mean	0.99	0.10	1.47	0.85	0.44	0.31
SEM	0.19	0.03	0.35	0.19	0.02	0.02

Table 3.1 Flexible model parameters. Mean and SEM across participants for all parameters in the Flexible model.

	$\bar{J}_{\text{total}}$	τ	α	β	γ
mean	0.88	0.10	1.54	1.23	0.56
SEM	0.17	0.04	0.36	0.42	0.29

Table 3.2 Minimizing Error model parameters. Mean and SEM across participants for all parameters in the Minimizing Error model.

4 Priority-modulated delay-period activity in visual cortex

When I was young, I said to God, 'God, tell me the mystery of the universe.' But God answered, 'That knowledge is for me alone.' So I said, 'God, tell me the mystery of the peanut.' Then God said, 'Well George, that's more nearly your size.' And he told me.

George Washington Carver

CABBAGE, n. A familiar kitchen-garden vegetable about as large and wise as a man's head

Ambrose Bierce, *The Devil's Dictionary*

4.1 Introduction

In previous chapters, we investigated the representation of priority and uncertainty in working memory. We showed that people maintain and use an accurate representation of uncertainty in working memory consistent with an optimal Bayesian observer (Ch. 1, 3). We showed that people use priority information to allocate resource to minimize their memory estimations (Ch. 2, 3), despite being incentivized in Ch. 3 to maximize points on a post-estimation wager. The goal of this chapter is to begin linking some of these computational concepts, priority, precision, and uncertainty, to neuroscience. In this chapter, we sppecifically investigate how priority is represented in the brain during a working memory delay.

Working memories are maintained through sustained, elevated delay-period activity. This was first demonstrated in the dorsolateral prefrontal cortex (Fuster & Alexander, 1971; Funahashi et al., 1989), but has since shown in sensory and motor areas (e.g. Curtis & D'Esposito, 2003; Postle, 2006; D'Esposito & Postle, 2015; D'Esposito, 2007; Y. Xu, 2017; Harrison & Tong, 2009). These data support a sensory recruitment theory, in which items are maintained by the same populations that encode them, rather than having to move or copy that information elsewhere. Frontal areas are then thought to play a role in top-down, goal-directed modulation of more sensory areas (Curtis & D'Esposito, 2003; Sreenivasan et al., 2014).

Because many frontal, parietal, and occipital areas of the brain are retinotopically organized, maintenance of information corresponds to sustained delay-period activity in the subpopulation sensitive to that items location. In a variety of tasks and brain areas, attention modulates this neural activity. This is typically through an increased activity of

neural populations tuned to the attended location, some research showing this effect more in frontoparietal than visual areas (Kastner et al., 1999; Buracas & Boynton, 2007; Gandhi et al., 1999; Gouws et al., 2014; Jerde et al., 2012; Serences & Yantis, 2007; Somers et al., 1999; Rahmati et al., 2018; Sprague et al., 2018; Saber et al., 2015; Nobre et al., 2004).

Because attention affects, physiologically, the activity of subpopulations during the delay and, behaviorally, the precision with which items are remembered (Bays & Husain, 2008; Emrich et al., 2017; Klyszejko et al., 2014, Ch. 2, Ch. 3), perhaps the magnitude of neural activity is a neural marker of precision. There is some support for this in neuroimaging and population coding models. Stimulus-independent, trial-to-trial fluctuations of neural activity in frontal areas have been shown to be positively correlated with fluctuations in performance (Sadaghiani, Hesselmann, Friston, & Kleinschmidt, 2010; Sapir, d'Avossa, McAvoy, Shulman, & Corbetta, 2005; Curtis, Rao, & D'Esposito, 2004; Rahmati et al., 2018). Population coding models implement memory precision of items through neural gain, which is able to account for human data on prioritization (Seung & Sompolinsky, 1993; Ma, Beck, Latham, & Pouget, 2006; Bays, 2014).

In this chapter, we investigate the resolution of this neural gain, asking if the amplitude of delay-period activity would reflect multiple levels of priority during a working memory delay. To test this, we used the same memory-guided saccade task used in Chapter 2 and 3. We hypothesized that priority is represented through the amplitude of the same populations that maintain that object's location. We defined ten visual, parietal, and frontal regions of interests (ROIs) known to be retinotopic and involved in working memory: V1, V2, V3, V3AB, IPS0, IPS1, IPS2, IPS3, iPCS, and sPCS. In each ROI, we tested our hypothesis by comparing the amplitudes of delay-period BOLD activity in the populations maintaining the location of each of the four VWM items.

4.2 Methods

4.2.1 Participants

Eleven people (5 males, mean age=31.9, SD=6.8, 5 authors) participated in this experiment. Everyone had normal or corrected-to-normal vision and no history of neurological disorders. Non-author participants were naive to the study hypotheses and were paid $30/hour. We obtained informed, written consent from all participants. The study was in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of New York University.

Experimental procedure Participants completed a memory-guided saccade task (Fig 4.1). The fixation symbol in this experiment was an encircled fixation cross, with four equally-spaced concentric arcs within each quadrant[1]. Each trial began with a 100 ms increase in the size of the outer circle of the fixation symbol. This was followed by a 700 ms endogenous precue which indicated the probe probability of each item. Probe probability was indicated through the number of illuminated arcs: all four arcs turned white in the quadrant corresponding to the 0.6 item, three arcs for the 0.3 item, two arcs for the 0.1 item, and none for the 0.0 stimulus. The precue was followed by a 100ms interstimulus interval, then by the items for 700 ms. The items were four white dots, one in each visual quadrant. Items were presented randomly between 9 and 10 degrees of visual angle (dva) from fixation. The location of the targets in polar coordinates were pseudo-randomly sampled from every 10 degrees, avoiding cardinal axes.

[1] The precue differed from previous experiment to ensure that visually-evoked responses of the precue were equivalent across priorities.

The item presentation was followed by a 10100 ms delay. A response cue appeared afterward, which was a white wedge around the quadrant of the fixation symbol corresponding to the target. Participants made a memory-guided saccade to the remembered dot location within the corresponding quadrant of the screen.

After the saccade, the actual dot location was presented as feedback and the participant made a corrective saccade to that location. After 800 ms, the feedback disappeared, participants returned their gaze to the central fixation cross, and a variable inter-trial interval began. The inter-trial interval was pseudorandomly drawn from the following three times: 8.8, 10.1 or 11.4 seconds. This was necessary to be able to disassociate different event-related activity in fMRI. Each participant completed one one-hour session consisting of 10-14 runs consisting of 12 trials each; they completed a total of 120-168 trials.

4.2.2 Experimental methods

4.2.3 Oculomotor methods

We recorded eye location data in the scanner at 1000 Hz (Eyelink 1000, SR Research, Ontario, Canada), beginning with a nine-point calibration and validation scheme. We processed the eye movement data using an in-house MATLAB function graphing toolbox (iEye). This toolbox transformed raw gaze positions into degrees of visual angle (dva), removed "extreme values" (defined as values that were larger than the screen size), removed artifacts due to blinks, smoothed gaze position with a Gaussian kernel with 5 ms SD, and computed the velocity at each time point. Saccades were defined as eye movements with the following criteria: velocity $\geq$ 30 dva/s, duration $\geq$ 8 ms, and amplitude $\geq$ 0.25 dva. For each trial, data were additionally drift corrected and calibrated to account for mea-

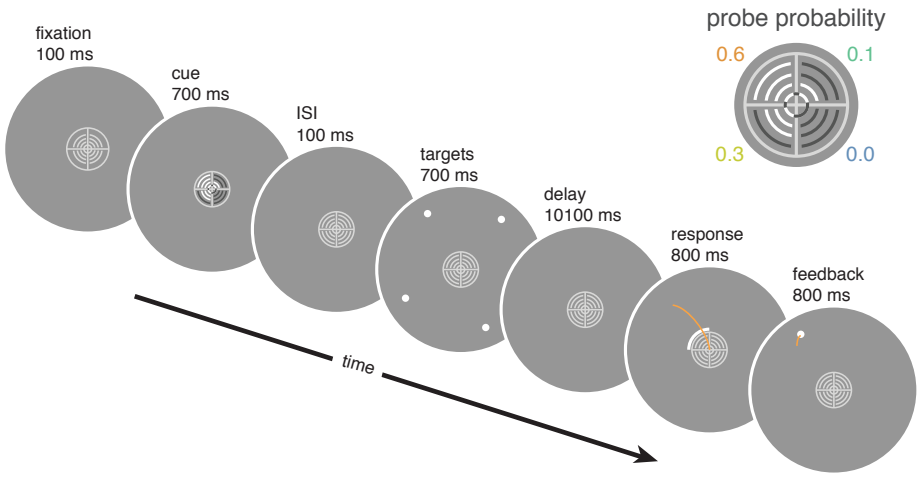

Figure 4.1 Trial sequence. Before targets were presented, participants viewed a precue that indicated the probe probabilities of the four targets by the number of arcs highlighted within the fixation symbol. After the delay, one item was probed for response when a white arc appeared at the outer edge of one quadrant of the fixation symbol. Participants made a memory-guided saccade to the remembered location of the target, then made a corrective saccade when the true target location was presented.

surement noise, such that the gaze position during known trial epochs (i.e., fixation and response period) were at the correct location. Trials were excluded if the participant was not fixating during the delay period or no saccades were found during the response epoch. This resulted in removing between 2% and 50% of trials per subject.

4.2.4 MRI acquisition

All functional MRI data and distortion scans were acquired on a 3T Siemens Prisma MRI system at the Center for Brain Imaging at New York University, using the CMRR Multi-Band Accelerated EPI Pulse Sequences (Release R015a; Moeller et al., 2010; Feinberg et al., 2010; J. Xu et al., 2013). To acquire the functional BOLD contrast images, we used the following settings: Multiband (MB) 2D GE-EPI with MB factor of 4, 56 2-mm interleaved slices with no gap, voxel size 2mm, field-of-view (FoV) 208 x 208 mm, no in-plane acceleration, repetition time (TR) 1300 ms, echo time (TE) 42 ms, flip angle 66 deg, Bandwidth: 1924 Hz/pixel (0.64 ms echo spacing), posterior-anterior phase encoding, with fat saturation and "brain" shim mode.

Distortion mapping scans, used to estimate the distortions present in the functional EPIs, were acquired with normal and reversed phase encoding after run 1, 3, 5, 7, and 9. We used a 2D SE-EPI with readout matching that of the GE-EPI and same number of slices, no slice acceleration, TE/TR: 45.6/3537 ms.

We used T1-weighted MP-RAGE scans (0.8 x 0.8 x 0.8 mm voxels, 256 x 240 mm FoV, TE/TR 2.24/2400 ms, 192 slices, bandwidth 210 Hz/Pixel, turbo factor 240, flip angle 8 deg, inversion non-selective (TI: 1060 ms)) for gray matter segmentation, cortical flattening, registration, and visualization for creating ROIs (details below).

4.2.5 fMRI processing

4.2.5.1 Preprocessing

During preprocessing of functional data, we align the brain across runs and account for run- and session-specific distortions, with the aim of minimizing spatial transformations. This allowed us to maximize signal to noise ratio and minimize smoothing, ensuring data remains as near as possible to its original resolution. All preprocessing was done in Analysis of Functional NeuroImages (AFNI, Cox, 1996).

First, we corrected functional images for intensity inhomogeneity induced by the high-density receive coil by dividing all images by a smoothed bias field (15 mm FWHM), computed as the ratio of signal in the receive field image acquired using the head coil to that acquired using the in-bore 'body' coil.

Next, we estimated distortion and motion-correction parameters. To minimize the effect of movement on the distortion correction (the distortion field depends on the exact position of the head in the main field), we collected multiple distortion scans throughout the experiment. Thus, every two functional runs flanked the distortion scans used to estimate these parameters. We refer to the functional-distortion-functional scan as a mini-session. For each mini-session, we used the distortion-correction procedure implemented in afni_proc.py to estimate parameters necessary to undistort and motion-correct functional images.

Then, we used the estimated distortion field, motion correction transform for each volume, and functional-to-anatomical coregistration simultaneously to render functional data from native acquisition space into unwarped, motion corrected, and coregistered anatomical

space for each participant at the same voxel size as data acquisition in a single transformation and resampling step. For each voxel on each run, we linearly detrended activation. We then computed percent signal change for each run.

4.2.5.2 Estimating delay-period activity

For each subject, we conducted a voxel-wise generalized linear model (GLM) to estimate each voxel's response to different trial events: precue, stimulus, delay, and response. The BOLD activity of a single voxel was predicted from a convolution of a canonical model of the hemodynamic impulse response function and a box-car regressor which had length equal to each trial event. We used one predictor each across all trials for the precue, stimulus, and response epochs. However, we defined a separate predictor for each trial for the delay period, so that we could have single trial estimates of delay-period activity. Additionally, there were predictors to account for motion and intercept for each run. This GLM was computed in AFNI.

4.2.6 Obtaining retinotopy

We used a recently developed population receptive field (pRF) mapping approach (Mackey, Winawer, & Curtis, 2017), which combines other pRF mapping approaches (Dumoulin & Wandell, 2008) with a more attentionally demanding task in order to map topographic areas in occipital, parietal, and frontal cortex. The methods are briefly summarized below; a more detailed description can be found in Mackey et al., 2017.

4.2.6.1 Behavioral task

Participants completed a difficult, covert attention task to ensure that the visual stimuli being presented were also being paid attention to. A trial consisted of three random dot kinematograms (RDK; Williams & Sekuler, 1984; Fig. 4.2A) within three adjacent rectangles. The participant indicated with a button press which of the two flanker rectangles contained dots moving in the same mean direction as the center rectangle. Importantly, these three rectangles, which formed a vertical or horizontal bar, swept across the visual field across the experiment, so that participants had to attend to the areas of the visual field that contained visual stimuli.

4.2.6.2 Estimating pRF

The predicted response amplitude for each voxel at time t, $\hat{r}(t)$, was modeled by the following equation[2]:

$$\hat{r}(t) = \gamma \left[\iint S(x,y) \mathcal{N}\left((x,y), \mathbf{I}\sigma\right) dxdy \right]^n, \tag{4.1}$$

in which S is a binary stimulus image (1s where the stimulus was presented and 0s otherwise), and $\mathcal{N}\left((x,y), \mathbf{I}\sigma^2\right)$ is a Normal distribution with mean (x,y) and variance $\mathbf{I}\sigma^2$, where $\mathbf{I}$ is a two-dimensional identity matrix. The parameters of this model are receptive field center (x,y), standard deviation σ, amplitude γ, and compressive spatial summation factor n. Parameters were fit with a course grid search over parameters, followed by

[2] A slight notation difference: In Mackey et al. (2017), γ is β. I switched it to decrease potential confusion with the βs obtained through the GLM

a local optimization method. Retinotopy estimates were used to define ROIs and estimate location-specific delay-period activity.

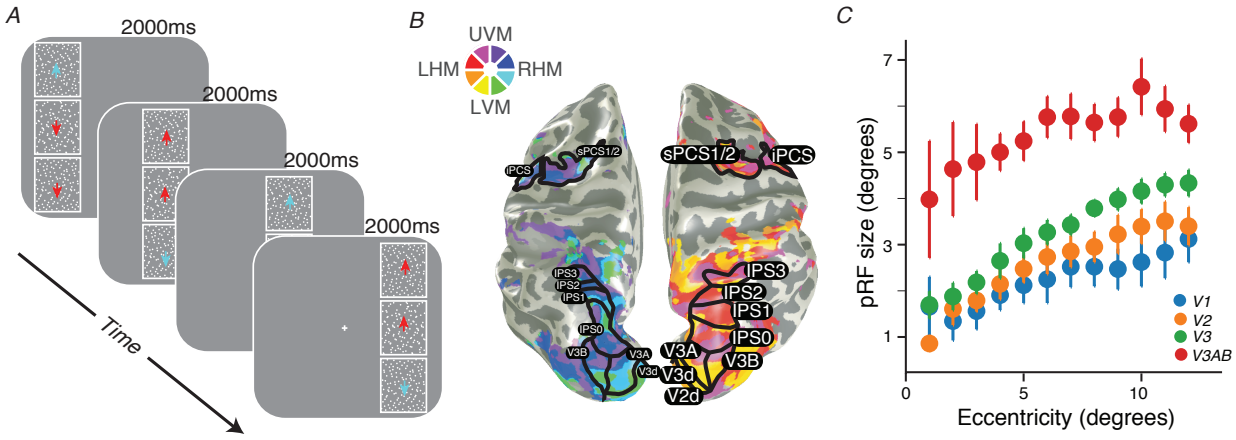

Figure 4.2 pRF-mapping schematics. *A*. The behavioral task used for pRF mapping. Participant indicated with a button press which of the two flanker rectangles contained dots moving in the same mean direction as the center rectangle. The rectangle configuration could sweep horizontally (as illustrated) or vertically. The dots within each rectangle moved orthogonal to the direction of the rectangle movement *B*. Example brain. The colors indicate the estimated polar angle center of each voxel's pRF, which are used to define ROIs. *C*. This plot, modified from Mackey et al. (2017), demonstrates that the estimated σ parameter increases across ROIs (as you move up the visual heirarchy and within an ROI (as you increase visual eccentricity). This provides evidence that pRF estimates reflect known characteristics of receptive fields of the visual system.

4.2.7 Defining ROIs

We used the estimated parameters acquired from the pRF mapping to define ROIs. Specifically, we used the estimates of the polar angle and eccentricity each voxel was most responsive to, as measured through the pRF model. We visualized flattened cortical surface representations of each subject's brain using AFNI and SUMA, and defined retinotopic maps based on standard conventions (Larsson & Heeger, 2006; Wandell, Dumoulin, & Brewer, 2007). We defined the following areas: V1, V2, V3, V3AB, IPS0, IPS1, IP2, IPS3, iPCS, and sPCS (Figure 4.2B).

In our analyses, we further restricted our ROIs based on pRF estimates. We excluded any voxels that did not have over 10% variance explained from the pRF model, and we excluded voxels with RF centers smaller than 4 dva or greater than 20 dva eccentricity from fixation.

Figure 4.2C is a modified figure from Mackey et al. (2017) demonstrating the sensible pRF estimates. Estimated receptive fields are larger with increasing eccentricity and larger cortical hierarchy.

4.2.8 Estimating item-specific delay-period BOLD activity

The goal of this study was to see if the priority of an item was reflected in the amplitude of the neural population encoding that item's location. This section summarizes how we obtained an estimate for each item's delay-period activity, using location estimates from the pRF mapping and delay-period activity estimates from the GLM. Theoretically, this measure weighs the delay-period activity of a voxel by its contribution to an item location.

For every item in every trial, we computed a pRF-weighted β,

$$\beta_{\text{pRF}} = \frac{1}{N} \sum_{i}^{N} w_i(x)\beta_i,$$

where $w_i(x)$ is the weight associated with the ith voxel at location x and β_i is the GLM-acquired delay-period β at voxel i. We define w_i as the receptive field of the voxel, which we model in accordance with the pRF models; each voxel's receptive field is represented as a non-normalized Gaussian with mean μ_i and variance σ_i^2.

$$w_i(x) = e^{\frac{-(x-\mu_i)^2}{2\sigma_i^2}}.$$

This formulation $w_i(x)$ results in voxels that are "more tuned" to an item's location to have a higher weight. For example, when the location is at the voxel's receptive field center, or when $x = \mu_i$, the weight $w_i = 1$. As the distance between x and μ_i increases, this weight decreases; the steepness with which it decreases is related to σ_i.

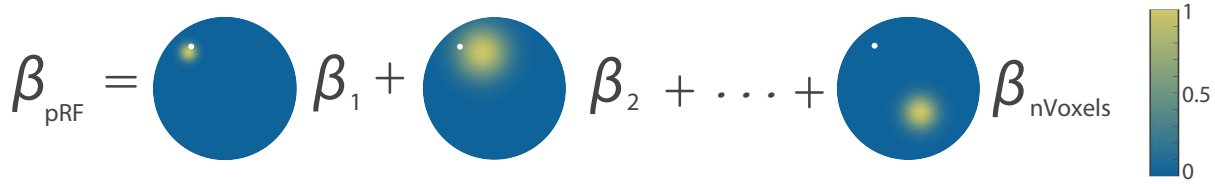

Figure 4.3 Schematic for calcuating item-specific delay-period amplitude, β_{pRF}. We weight the GLM-obtained estimate of delay-period activity for voxel i, β_i, by its sensitivity to the location of the current item (as estimated by the pRF model). The heat plot illustrates the estimated receptive field over the entire stimulus display, a nonnormalized Gaussian distribution with a mode equivalent to weight = 1. The corresponding weight for each voxel is the value at the item's location, illustrated by the white dot. Note that this schematic implies that β_{pRF} is a sum, but it is actually a mean.

For each ROI, we conducted a repeated-measures ANOVA (rmANOVA) with priority as the within-subjects factor and β_{pRF} as the dependent variable.

4.3 Results

4.3.1 Behavioral results

Error decreased with increasing priority. This effect was marginally significant when corrected for violating sphericity ($F(1.2, 12.4) = 3.53, p = 0.08, \eta^2 = 0.16$).

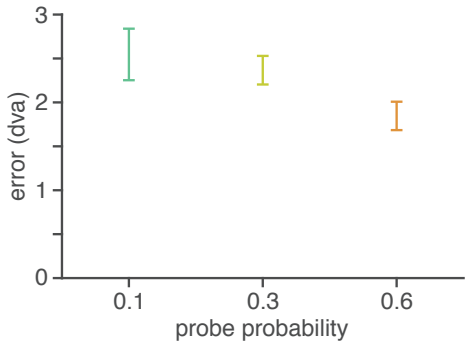

Figure 4.4 Behavioral results. Final saccade error ($M \pm SEM$ across participants) decreases with increasing priority.

4.3.2 Neuroimaging results

For each participant and ROI, I averaged the β_{pRF}s corresponding to each priority across trials. For each ROI, I conducted a repeated-measures ANOVA with the priority condition as the within-subject variable and β_{pRF} as the dependent variable. Because we assume the data from each ROI is largely independent, we chose not to do a multiple correction to account for the multiple ANOVAs across ROIs[3]. Within each ROI, all post hoc tests were adjusted for multiple comparisons, and we will report significance values before and after

[3] This would not affect the results or interpretation

a Bonferroni correction. We found an effect of priority on the delay-period activity of neural populations in visual areas, but not in parietal or frontal areas. All ANOVAs met the assumption of sphericity, as tested through Mauchly's test for sphericity.

4.3.2.1 V1

For visual area V1, there was a significant effect of priority on delay-period BOLD activity, $F(3, 30) = 18.04, p = 7 \times 10^{-7}$, first column of Figure 4.5. Post hoc tests indicate that the high priority ($M \pm SEM$: -0.01 ± 0.01) was significantly higher than 0.3 ($-0.03 \pm 0.01, t(30) = -3.90, p = .002$ (Bonferroni corrected $p = .02$)), 0.1 ($-0.03 \pm 0.01, t(30) = -4.52, p = .0005$ (.003)), and 0.0 ($-0.04 \pm 0.01, t(30) = -7.28, p = 2 \times 10^{-7}$ (1×10^{-6})). The medium priority activity was significantly higher than 0.0 ($t(30) = -3.38, p = 0.01$ (.06)). The low priority activity was higher than 0.0, but the significance did not survive multiple comparisons ($t(30) = -2.76, p = .05$ (.3)).

4.3.2.2 V2

For visual area V2, there was a significant effect of priority on delay ($F(3, 30) = 18.60, p = 5 \times 10^{-7}$, first column of Figure 4.5). Post hoc tests indicated that the high priority (0.01 ± 0.01) was significantly higher than 0.3 ($-0.01 \pm 0.01, t(30) = -3.72p = .004$ (.03)), 0.1 ($-0.02 \pm 0.01, t(30) = -6.12p = 6 \times 10^{-6}$ (3×10^{-5})), and 0.0 ($-0.03 \pm 0.01, t(30) = -6.73p = 1 \times 10^{-6}$ ($6e-6$)). The medium priority activity was significantly higher than 0.0 ($t(30) = -3.00p = .03$ (.16)).

4.3.2.3 V3

For visual area V3, there was a significant effect of priority on delay ($F(3, 30) = 12.5, p = 2 \times 10^{-5}$, third column of Figure 4.5). Post hoc tests indicated that the high priority (0.01 ± 0.01) was significantly higher than 0.3 ($-0.01 \pm 0.01, t(30) = -3.2, p = .02$ (.1)), 0.1 ($-0.01 \pm 0.01, t(30) = -4.2, p = .001$ (.006)), and 0.0 ($-0.02 \pm 0.01, t(30) = -5.9, p = 1 \times 10^{-5} (6 \times 10^{-5})$). Additionally, the medium priority activity was significantly higher than 0.0 ($t(30) = -2.8, p = .04$ (.3)). Note that some of these analyses did not survive multiple comparisons.

4.3.2.4 V3AB

For visual area V3AB, there was a significant effect of priority on delay ($F(3, 30) = 5.87, p = .003$, fourth column of Figure 4.5). Post hoc tests indicated that the high priority (0.06 ± 0.01) was significantly higher than 0.3 ($0.04 \pm 0.01, t(30) = -3.04, p = 0.02$ (.14)), 0.1 ($0.04 \pm 0.01, t(30) = -3.03, p = 0.02$(.15)), and 0.0 ($0.03 \pm 0.01, t(30) = -3.91, p = 0.002$(.02)). Only the comparison between 0.6 and 0.0 survived multiple comparisons.

4.3.2.5 IPS0-3, iPCS, sPCS

In all parietal and frontal ROIs, there was no effect of priority on delay-period activity, as measured through the β_{pRF}, $p > .05$, Figure 4.6.

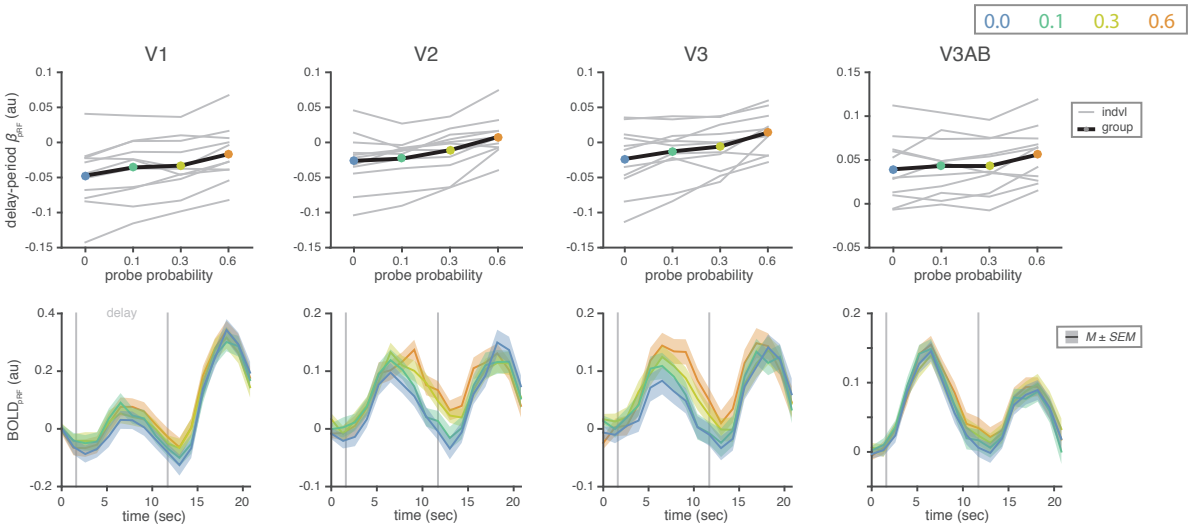

Figure 4.5 fMRI results in visual areas. Each column corresponds to a different ROI in visual cortex. The first row illustrates the β_{pRF}s as a function of priority. Individual participant averages are shown in grey lines, and the group average is shown in black. There is an increase in β_{pRF} with priority. The bottom row illustrates the trial BOLD signal ($M \pm SEM$ across trials), pRF-weighted in the same fashion as the β_{pRF}. The BOLD_{pRF} traces diverge based on priority after the evoked response associated with the cue onset.

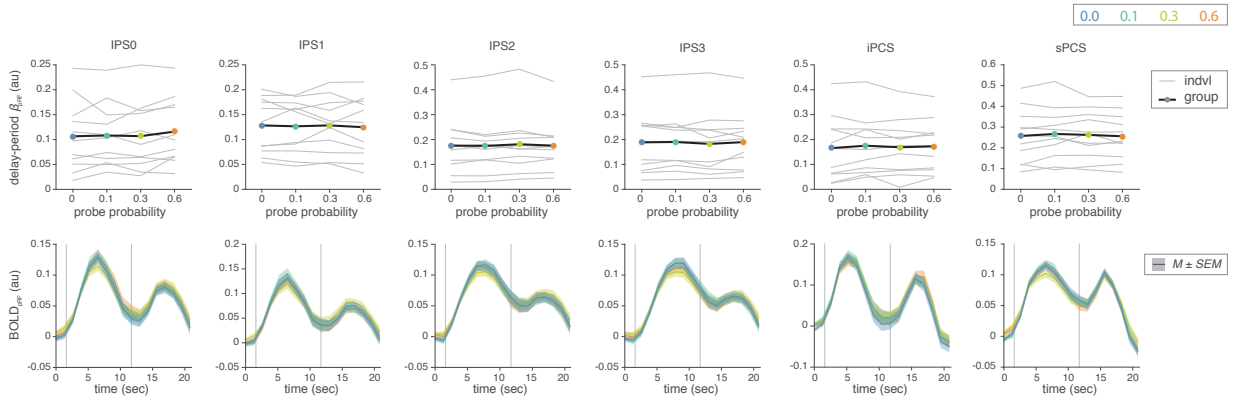

Figure 4.6 fMRI results in parietal and frontal areas. Each column corresponds to a different ROI in parietal or frontal areas. The first row illustrates the β_{pRF}s as a function of priority. Individual participant averages are shown in grey lines, and the group average is shown in black. The bottom row illustrates the trial BOLD signal ($M \pm SEM$ across trials), pRF-weighted in the same fashion as the β_{pRF}. There is no qualitative effect of priority on either the β_{pRF} or BOLD_{pRF}.

4.4 Discussion

In this study, we asked how priority was represented in working memory. Specifically, we hypothesized that an item's priority would be represented through the gain of the same neural populations that encode its location. To test this, we collected fMRI BOLD activity while participants completed a spatial delayed estimation task and analyzed the delay period activity.

Behaviorally, error decreased with increasing priority, bearing resemblance to the results in Ch. 2 and 3. We thus suspect the marginal significance was a result of low statistical power; we had substantially less trials than in previous studies.

Neurally, we found that delay-period BOLD activity tracked priority in visual areas V1, V2, V3, V3AB. These results are consistent with a subset of the literature investigating the effects of set size on neural activity. Items are equally behaviorally relevant in set size studies and thus the priority of each item is $\frac{1}{N}$, where N is the set size. With higher set size, the priority of each item is lower. We can thus consider set size as an indirect priority manipulation. Increasing set size (i.e., decreasing priority per item) is associated with a decrease in target reconstructions (Sprague, Ester, & Serences, 2014, 2016), activity in monkey striate cortex (Landman, Spekreijse, & Lamme, 2003), and classifier accuracy (Emrich, Riggall, Larocque, & Postle, 2013).

Priority, however, did not modulate delay activity in parietal and frontal areas. Across all levels of priority, the amount of delay-period activity in populations tuned to the items location was remarkably equivalent. This is not to imply that there is no maintenance, there is clear delay-period activity (bottom row of Fig. 4.6). The results of this section

might corroborate the larger subsection of the set size literature, showing that in frontoparietal areas neural activity seems to increase with increasing set size (Todd & Marois, 2004; Y. Xu & Chun, 2006; Sprague et al., 2014; Emrich et al., 2013; Braver et al., 1997). Perhaps, within some limit, the addition of every item in memory comes with a somewhat fixed amount of activity and thus higher set size would necessitate higher activity.

However, this result seems somewhat inconsistent with the notion of frontal and parietal regions maintaining more goal-oriented information than visual areas and previous research has shown strong attentional modulation in higher-level areas (Sreenivasan et al., 2014; Curtis & D'Esposito, 2003; Kastner et al., 1999; Gouws et al., 2014; Serences & Yantis, 2007; Somers et al., 1999). Why might we have found this result? Perhaps priority was being used to create priority maps in visual areas, and did not need to be maintained afterward. There were 1.5 seconds between the precue onset and delay onset, which allowed ample time for frontoparietal areas to modulate the activity of sensory areas. Once it has successfully communicated goal-directed information, it may not need to maintain it anymore. On the other hand, frontoparietal voxels may not be sensitive enough to location for this analysis to work. The estimated σs in frontoparietal regions are extremely large (and are noisier estimates). If voxels are not sensitive to location, then the β_{pRF} will not differ across items. Finally, priority could be maintained, but not through the amplitude of the neural population. Perhaps the amount of activity in these regions is related to the amount of top-down information. Each subpopulation would be holding a number (i.e., the priority), and thus may have different patterns of activity with the same overall amplitude.

4.4.1 Limitations and future directions

We all know Spidermans's Uncle Ben's famous words: "with every methodological choice comes the responsibility of dealing with the limitations associated with it."[4] We think it is important to state what our assumptions are and how, if at all, violations of these assumptions could affect the results.

One of the largest assumption we make in our analyses is that pRF estimates are stable across time, independent of stimulus and priority configuration There is strong empirical evidence that this assumption is violated. There are studies demonstrating that priority, through attention or set size manipulations, can affect pRFs by doing a combination of changing the center, size, and gain (de Haas, Schwarzkopf, Anderson, & Rees, 2014; Kay, Weiner, & Grill-Spector, 2015; Klein, Harvey, & Dumoulin, 2014; Sheremata & Silver, 2015; Vo, Sprague, & Serences, 2017). The effects of attention may differentially effect different brain areas or even hemispheres (Klein et al., 2014; Sheremata & Silver, 2015). Based on this result, it is possible that on every trial, the configurations of stimuli and, more importantly, the priority associated with each, would sculpt the pRFs in different ways. It is possible that our assumption leads to poorer estimates of what the delay-period activity for each item would actually be. However, we do not believe the assumption could cause the result we observed. Because we randomly picked the location of objects within each quadrant and the priorities associated with each, any trial-specific fluctuations in pRFs would averaged away across trials.

Additionally, we do not include two parameters from our pRF model of neural activity (Eq. 4.1) when calculating the pRF-weighted delay-period activity: an amplitude γ and

[4] This is definitely not what he said.

compressive spatial summation factor n. The former accounts for differences in voxel activity. We chose to exclude this because the GLM estimates also account for this, and we didn't want to "double up" on the same parameter. The latter parameter accounts for the fact that stimulus-evoked responses grow nonlinearly with stimulus intensity (Kay, Winawer, Mezer, & Wandell, 2013). Because for each trial, each estimate of item-specific delay-period activity is summed across all voxels, the relative contribution with and without this parameter would not change. However, the weighted average across voxels could, in theory, be affected by this parameter. To investigate this, I plan to recalculate the estimate of item-specific delay-period activity, β_{pRF}, with these parameters.

4.4.2 Conclusions

We show first evidence of a truly graded representation of an item's priority in the same neural populations that maintain its location. We find this only in visual cortex, not in higher-level areas, despite clear delay-period activity in these areas. These results provide evidence that the distribution of WM resource according to priority sculpts the relative gains of neural populations in visual areas that encode items. Furthermore, our results demonstrate the existence of different representations across the processing heirarchy, supporting their different roles. More generally, our result contribute evidence in favor of sensory areas storing not only a point estimate of individual items, but some aspect related to memory fidelity. We think this result is promising, and hope it serves as an exciting step toward finding and disambiguating the neural mechanisms responsible for processing and maintaining computational concepts such as priority, uncertainty, and precision.

5 Conclusions

I'm going to the bathroom to read.

Elvis Presley

5.0.1 Summary of dissertation

In Chapter 1, we asked if people could maintain and use a representation of uncertainty in working memory. We tested this question by seeing if they used uncertainty in a four-item orientation change detection task. We factorially compared models with different hypotheses about how items are encoded into memory, how observers believed their memory was encoded, and their decision rule We found that an observer who correctly assumed that their precision varied on an item-to-item and trial-to-trial basis fit the data best. A model who assumed information was optimally combined performed slightly better than one that assumed people didn't. This chapter provides evidence of probabilistic computation in working memory.

In Chapter 2, we asked whether people used priority information when allocating resource across items. We tested this question by using a four-item spatial delayed-estimation task, where each item had a different priority. We found behaviorally that error decreased with increasing priority, indicating that people allocated resource according to priority. We used computational modeling to investigate participants' allocation strategies, and found that they were best fit by a model which assumes people are allocating resource to minimize a function of estimation error.

In Chapter 3, we wanted to test the generalizability of the previous two chapters. We investigated this by collecting a four-item memory-guided saccade task, with a post-estimation wager. We found the following three behavioral results: error decreased with increasing priority, indicating people were allocating resource according to priority; circle size decreased with increasing priority, indicating people had lower uncertainty for higher priority items; and these two values were correlated, indicating that people had an item-

specific representation of uncertainty that reflected the precision with which items were remembered. Computationally, we found that people set circle size consistent with an optimal model. Additionally, we found that participants allocated resource in order to minimize a function of estimation error, despite being incentivized to maximize points in the post-estimation wager. These results suggest that our ability to use uncertainty information is flexible across tasks, but allocation according to priority is less flexible. Future research should investigate what affects the flexibility of resource allocation. Is it truly flexible?

In Chapter 4, we asked how priority was represented in the brain during a working memory delay. We tested this question by collecting BOLD activity while participants completed the same four-item memory-guided saccade task. We hypothesized that priority would be represented in the delay-period amplitude of populations maintaining each item's location. We found this effect in visual areas, indicating that priority, and perhaps precision, is represented in the same populations that encode the item's location. However, we did not find this effect in parietal or frontal areas. Further analyses must be done to investigate whether and how priority is maintained in these areas. Either way, these results provide some insight into how priority is represented in the brain, and highlight the different representations and functional roles of sensory and higher level areas.

Overall, this thesis demonstrates two ways people compensate for having a limited-resource working memory. One way is to know our memory fidelity, and make decisions based on that knowledge. The another is to remember more important things better.

5.0.2 Relation to broader literature

There are, of course, other strategies people are using to facilitate visual working memory. There is evidence that people use statistical regularities of stimuli to remember items more precisely (Brady & Tenenbaum, 2013; Brady, Konkle, & Alvarez, 2009; Victor & Conte, 2004). People can incorporate prior knowledge into their representations (Honig, Ma, & Fougnie, 2018). People may use various compression strategies that, while potentially inducing a bias, facilitate memory for more items (Nassar et al., 2018; Bays, Catalao, & Husain, 2009). (Although it could be argued, however, that these compression strategies are not intentional, but a result of attraction and repulsion dynamics of different neural populations (Almeida, Barbosa, & Compte, 2015).)

It is possible that the specific methodological choices we made could have affected our results. For example, location is a special feature, indicated by the retinotopic structure of the brain. Intuitively, it may be special because location is a necessary feature for all objects or because we bind different features to one object by their shared location. There is some computational evidence that location is the linking feature between all other features of an object (Schneegans & Bays, 2017). Would our results on priority generalize to other features? Second, some argue that the strong delay-period activity shown in many classical working memory tasks, which use oculomotor responses, could be driven by motor planning, rather than working memory (e.g. Lundqvist, Herman, & Miller, 2018). Would a different response modality affect our results in Chapter 4? Last, because we use a precue, our effects in Chapter 2, 3, and 4 are a result of attentional allocation to perceptual stimuli, rather than a reallocation of resource to items already in working memory (e.g. Griffin & Nobre, 2003). How would the behavioral and neural results change with a retrocue

paradigm? Investigating how subtle changes in methodologies change results could help illuminate how working memory information is processed and maintained for later use.

Finally, throughout this dissertation, I assume working memory is a stochastic, continuous resource that can be shared across items. While there is considerable evidence within this dissertation that memory fidelity fluctuates on an item-to-item basis (see model comparison results of Ch. 1 and behavioral correlation in Ch. 4), we do not make any claims about whether the VP model is the best computational model to capture the data. In fact, we recognize that the VP model is a tool to understand how information is processed and used in working memory, but that it does not provide an explanation for how this process works in the brain. The VP model's unconstrained assumptions regarding resource allocation were particularly appealing for Chapter 2 and 3, but alternative models such as the Interference (Oberauer & Lin, 2017), Slots+Resources, or Slots+Averaging (Zhang & Luck, 2008) models may additionally provide a good fit to the data. Future research could investigate these models.

5.0.3 Conclusion

In everyday life, we are bombarded with information constantly, and we have to decide what to look at, pay attention to, and remember. The studies within this dissertation provide evidence that people behave rationally in relation to their working memory limits. First, we remember more important items better, in performance-related way. Second, we are aware of and account for our item-specific memory fidelity when making working memory decisions. Finally, the priority, and perhaps precision, of an item is represented through the amplitude of delay-period activity in the same sensory populations that encode its location. The results of this dissertation demonstrate different ways in which we

use task-relevant information to adjust how we encode information and make decisions in working memory.

References

Acerbi, L., & Ma, J. W. (2017). *Practical bayesian optimization for model fitting with bayesian adaptive direct search* (Vol. 30). Advances in Neural Information Processing Systems.

Acerbi, L., Ma, W. J., & Vijayakumar, S. (2014). A framework for testing identifiability of bayesian models of perception. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 27* (pp. 1026–1034). Curran Associates, Inc.

Adam, S. K. C., Robison, K. M., & Vogel, K. E. (2018). Contralateral delay activity tracks fluctuations in working memory performance. *Journal of cognitive neuroscience*, 1–12. doi: 10.1162/jocn_a_01233

Adam, S. K. C., & Vogel, K. E. (2017, Jul 1). Confident failures: Lapses of working memory reveal a metacognitive blind spot. *Attention, perception & psychophysics*, *79*(5), 1506–1523. doi: 10.3758/s13414-017-1331-8

Alais, D., & Burr, D. (2004, Feb 3). The ventriloquist effect results from near-optimal bimodal integration. *Current biology: CB*, *14*(3), 257–262. doi: 10.1016/j.cub.2004.01.029

Almeida, R., Barbosa, J., & Compte, A. (2015, Sep 15). Neural circuit basis of visuospatial working memory precision: a computational and behavioral study. *Journal of neurophysiology*, *114*(3), 1806–1818. doi: 10.1152/jn.00362.2015

Alvarez, A. G., & Cavanagh, P. (2004, Feb 1). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science*, *15*(2), 106–111. doi: 10.1111/j.0963-7214.2004.01502006.x

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: the "oblique effect" in man and animals. *Psychological Bulletin*, *78*(4), 266–278.

Atkinson, L. A., Baddeley, D. A., & Allen, J. R. (2018, Jul 1). Remember some or remember all? ageing and strategy effects in visual working memory. *Quarterly journal of experimental psychology (2006)*, *71*(7), 1561–1573. doi: 10.1080/17470218.2017.1341537

Baddeley, A. (2003, Oct 1). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839. doi: 10.1038/nrn1201

Baddeley, A., & Hitch, G. (1974). Working memory. *The psychology of learning and motivation*, *8*, 47–89.

Barthelme, S., & Mamassian, P. (2010, Nov 30). Flexible mechanisms underlie the evaluation of visual confidence. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(48), 20834–20839. doi: 10.1073/pnas.1007704107

Bays, M. P. (2014). Noise in neural populations accounts for errors in working memory. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *34*(10), 3632–3645.

Bays, M. P., Catalao, G. R. F., & Husain, M. (2009, Sep 9). The precision of visual working memory is set by allocation of a shared resource. *Journal of vision*, *9*(10), 7.1–711. doi: 10.1167/9.10.7

Bays, M. P., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854.

Beck, M. J., Ma, J. W., Pitkow, X., Latham, E. P., & Pouget, A. (2012, Apr 12). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron*, *74*(1), 30–39. doi: 10.1016/j.neuron.2012.03.016

Bengson, J. J., & Luck, J. S. (2016, Feb 1). Effects of strategy on visual working memory capacity. *Psychonomic bulletin & review*, *23*(1), 265–270. doi: 10.3758/s13423-015-0891-7

Bisley, W. J., & Goldberg, E. M. (2010). Attention, intention, and priority in the parietal lobe. *Annual review of neuroscience*, *33*, 1–21. doi: 10.1146/annurev-neuro-060909-152823

Bona, S., & Silvanto, J. (2014, Mar 24). Accuracy and confidence of visual short-term memory do not go hand-in-hand: behavioral and neural dissociations. *PloS one*, *9*(3). doi: 10.1371/journal.pone.0090808

Brady, F. T., Konkle, T., & Alvarez, A. G. (2009, Nov 1). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of experimental psychology. General*, *138*(4), 487–502. doi: 10.1037/a0016797

Brady, F. T., Konkle, T., Alvarez, A. G., & Oliva, A. (2008, Sep 23). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14325–14329. doi: 10.1073/pnas.0803390105

Brady, F. T., Konkle, T., Gill, J., Oliva, A., & Alvarez, A. G. (2013, Jun 29). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological science*, *24*(6), 981–990. doi: 10.1177/0956797612465439

Brady, F. T., & Tenenbaum, B. J. (2013, Jan 10). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological review*, *120*(1), 85–109. doi: 10.1037/a0030779

Braver, S. T., Cohen, D. J., Nystrom, E. L., Jonides, J., Smith, E. E., & Noll, C. D. (1997, Jan 1). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*, *5*(1), 49–62. doi: 10.1006/nimg.1996.0247

Buracas, T. G., & Boynton, M. G. (2007, Jan 3). The effect of spatial attention on contrast response functions in human visual cortex. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *27*(1), 93–97. doi: 10.1523/JNEUROSCI.3162-06.2007

Chen, L. L., Chen, M. Y., Zhou, W., & Mustain, D. W. (2014). Monetary reward speeds up voluntary saccades. *Frontiers in integrative neuroscience*, *8*. doi: 10.3389/fnint.2014.00048

Conway, A. A. R., Kane, J. M., & Engle, W. R. (2003, Dec 1). Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences*, *7*(12), 547–552. doi: 10.1016/j.tics.2003.10.005

Cowan, N. (2001, Feb 1). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and brain sciences*, *24*, 87–114; discussion 114. doi: 10.1017/S0140525X01003922

Cox, W. R. (1996, Jun 1). Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and biomedical research, an international journal*, *29*(3), 162–173.

Curtis, E. C., & D'Esposito, M. (2003, Sep 1). Persistent activity in the prefrontal cortex during working memory. *Trends in cognitive sciences*, *7*(9), 415–423. doi: 10.1016/S1364-6613(03)00197-9

Curtis, E. C., Rao, Y. V., & D'Esposito, M. (2004, Apr 21). Maintenance of spatial and motor codes during oculomotor delayed response tasks. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *24*(16), 3944–3952. doi: 10.1523/JNEUROSCI.5640-03.2004

de Haas, B., Schwarzkopf, S. D., Anderson, J. E., & Rees, G. (2014, Jan 20). Perceptual load affects spatial tuning of neuronal populations in human early visual cortex. *Current biology: CB*, *24*(2), R66–R67. doi: 10.1016/j.cub.2013.11.061

De Silva, N., & Ma, J. W. (2017). Optimal allocation of attentional resource to multiple items with unequal relevance. *arXiv*.

D'Esposito, M. (2007, May 29). From cognitive to neural models of working memory. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *362*(1481), 761–772. doi: 10.1098/rstb.2007.2086

D'Esposito, M., & Postle, R. B. (2015, Jan 3). The cognitive neuroscience of working memory. *Annual review of psychology*, *66*, 115–142. doi: 10.1146/annurev-psych-010814-015031

Dube, B., Emrich, M. S., & Al-Aidroos, N. (2017). More than a filter: Feature-based attention regulates the distribution of visual working memory resources. *Journal of experimental psychology. Human perception and performance*, *43*(10), 1843–1854.

Dumoulin, O. S., & Wandell, A. B. (2008, Jan 15). Population receptive field estimates in human visual cortex. *NeuroImage*(2), 647–60. doi: 10.1016/j.neuroimage.2007.09.034

Elmore, C. L., Ma, J. W., Magnotti, F. J., Leising, J. K., Passaro, D. A., Katz, S. J., & Wright, A. A. (2011, Jun 7). Visual short-term memory compared in rhesus monkeys and humans. *Current biology: CB*, *21*(11), 975–979. doi: 10.1016/j.cub.2011.04.031

Emrich, M. S., Lockhart, A. H., & Al-Aidroos, N. (2017). Attention mediates the flexible allocation of visual working memory resources. *Journal of experimental psychology. Human perception and performance*. doi: 10.1037/xhp0000398

Emrich, M. S., Riggall, C. A., Larocque, J. J., & Postle, R. B. (2013, Apr 10). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *The Journal of neuroscience: the of-*

ficial journal of the Society for Neuroscience, *33*(15), 6516–6523. doi: 10.1523/JNEUROSCI.5732-12.2013

Engle, W. R., Kane, J. M., & Tuholski, W. S. (1998). Models of working memory: Mechanisms of active maintenance and executive control. In A. Miyake & P. Shah (Eds.), (pp. 102–134). Cambridge University Press. doi: 10.1017/CBO9781139174909.007

Ernst, O. M., & Banks, S. M. (2002, Jan 24). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. doi: 10.1038/415429a

Feinberg, A. D., Moeller, S., Smith, M. S., Auerbach, E., Ramanna, S., Gunther, M., ... Yacoub, E. (2010, Dec 20). Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging. *PloS one*, *5*(12). doi: 10.1371/journal.pone.0015710

Fougnie, D., Asplund, L. C., & Marois, R. (2010, Oct 22). What are the units of storage in visual working memory? *Journal of vision*, *10*(12). doi: 10.1167/10.12.27

Fougnie, D., Cormiea, M. S., Kanabar, A., & Alvarez, A. G. (2016). Strategic trade-offs between quantity and quality in working memory. *J of exp psychol: HPP*, *42*(8), 1231–1240.

Fougnie, D., Suchow, W. J., & Alvarez, A. G. (2012). Variability in the quality of visual working memory. *Nature communications*, *3*.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010, Oct 1). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychonomic bulletin & review*, *17*(5), 673–679. doi: 10.3758/17.5.673

Funahashi, S., Bruce, J. C., & Goldman-Rakic, S. P. (1989, Feb 1). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of neurophysiology*, *61*(2), 331–349. doi: 10.1152/jn.1989.61.2.331

Furmanski, S. C., & Engel, A. S. (2000). An oblique effect in human primary visual cortex. *Nature Neuroscience*, *3*.

Fuster, M. J., & Alexander, E. G. (1971, Aug 13). Neuron activity related to short-term memory. *Science (New York, N.Y.)*, *173*(3997), 652–654. doi: 10.1126/science.173.3997.652

Gandhi, P. S., Heeger, J. D., & Boynton, M. G. (1999, Mar 16). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(6), 3314–3319. doi: 10.1073/pnas.96

.6.3314

Girshick, R. A., Landy, S. M., & Simoncelli, P. E. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*.

Goldberg, D. (1988). *Genetic algorithms in search, optimization, and machine learning, addison wesley, reading, ma.* Addison Wesley.

Gorgoraptis, N., Catalao, G. R. F., Bays, M. P., & Husain, M. (2011, Jun 8). Dynamic updating of working memory resources for visual objects. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *31*, 8502–8511. doi: 10.1523/JNEUROSCI.0208-11.2011

Gouws, D. A., Alvarez, I., Watson, M. D., Uesaki, M., Rogers, J., & Morland, B. A. (2014). On the role of suppression in spatial attention: evidence from negative bold in human subcortical and cortical structures. *J Neurosci.*, *34*(31), 10347–60.

Graf, W. E., Warren, A. P., & Maloney, T. L. (2005). Explicit estimation of visual uncertainty in human motion processing. *Vision research*, *45*(22), 3050–3059.

Griffin, C. I., & Nobre, C. A. (2003, Nov 15). Orienting attention to locations in internal representations. *Journal of cognitive neuroscience*, *15*(8), 1176–1194. doi: 10.1162/089892903322598139

Hanson, N., Niederberger, P. A. S., Guzella, L., & Koumoutsakos, P. (2008). A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion. *IEEE Transactions on Evolutionary Computation*, *13*(1), 180–197.

Harris, M. C., & Wolpert, M. D. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780–784.

Harrison, A. S., & Tong, F. (2009, Apr 2). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. doi: 10.1038/nature07832

Honig, M., Ma, J. W., & Fougnie, D. (2018). Humans incorporate trial-to-trial working memory uncertainty into rewarded decisions. *bioRXiv*.

Huber, P. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, *35*(1), 73–101. doi: 10.1214/aoms/1177703732

Hurvich, M. C., & Tsai, L. C. (1987). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307.

Jerde, A. T., Merriam, P. E., Riggall, C. A., Hedges, H. J., & Curtis, E. C. (2012, Nov 28). Prioritized maps of space in human frontoparietal cortex. *J Neurosci*, *32*(48), 17382–17390. doi: 10.1523/JNEUROSCI.3810-12.2012

Just, A. M., & Carpenter, A. P. (1992, Jan 1). A capacity theory of comprehension: individual differences in working memory. *Psychological review*, *99*(1), 122–149. doi: 10.1037/0033-295X.99.1.122

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, *47*, 263–291.

Kastner, S., Pinsk, A. M., De Weerd, P., Desimone, R., & Ungerleider, G. L. (1999, Apr 1). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–761. doi: 10.1016/S0896-6273(00)80734-5

Kato, M., Miyashita, N., Hikosaka, O., Matsumura, M., Usui, S., & Kori, A. (1995). Eye movements in monkeys with local dopamine depletion in the caudate nucleus. i. deficits in spontaneous saccades. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *15*(1), 912–927.

Kay, N. K., Weiner, S. K., & Grill-Spector, K. (2015, Mar 2). Attention reduces spatial uncertainty in human ventral temporal cortex. *Current biology: CB*, *25*(5), 595–600. doi: 10.1016/j.cub.2014.12.050

Kay, N. K., Winawer, J., Mezer, A., & Wandell, A. B. (2013). Compressive spatial summation in human visual cortex. *Journal of neurophysiology*, *110*(2). doi: 10.1152/jn.00105.2013

Keshvari, S., Berg, d. R. v., & Ma, J. W. (2012). Probabilistic computation in human perception under variability in encoding precision. *PLoS ONE*, *7*.

Keshvari, S., van den Berg, R., & Ma, J. W. (2013, Feb 28). No evidence for an item limit in change detection. *PLoS computational biology*, *9*. doi: 10.1371/journal.pcbi.1002927

Klein, P. B., Harvey, M. B., & Dumoulin, O. S. (2014, Oct 1). Attraction of position preference by spatial attention throughout human visual cortex. *Neuron*, *84*(1), 227–237. doi: 10.1016/j.neuron.2014.08.047

Klyszejko, Z., Rahmati, M., & Curtis, E. C. (2014). Attentional priority determines working memory precision. *Vision research*, *105*, 70–76. doi: 10.1016/j.visres.2014.09.002

Knill, C. D., & Pouget, A. (2004, Dec 1). The bayesian brain: the role of uncertainty in neural coding and computation. *Trends in neurosciences*, *27*(12), 712–719. doi: 10.1016/j.tins.2004.10.007

Konkle, T., Brady, F. T., Alvarez, A. G., & Oliva, A. (2010, Aug 1). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of experimental psychology. General*, *139*(3), 558–578. doi: 10.1037/a0019165

Kording, K. (2007, Oct 26). Decision theory: what "should" the nervous system do? *Science (New York, N.Y.)*, *318*(5850), 606–610. doi: 10.1126/science.1142998

Landman, R., Spekreijse, H., & Lamme, F. V. A. (2003, Aug 15). Set size effects in the macaque striate cortex. *Journal of cognitive neuroscience*, *15*(6), 873–882. doi: 10.1162/089892903322370799

Larsson, J., & Heeger, J. D. (2006, Dec 20). Two retinotopic visual areas in human lateral occipital cortex. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *26*(51), 13128–13142. doi: 10.1523/JNEUROSCI.1657-06.2006

Lee, E.-Y., Cowan, N., Vogel, K. E., Rolan, T., Valle-Inclán, F., & Hackley, A. S. (2010, Sep 5). Visual working memory deficits in patients with parkinson's disease are due to both reduced storage capacity and impaired ability to filter out irrelevant information. *Brain: a journal of neurology*, *133*(9), 2677–2689. doi: 10.1093/brain/awq197

Lee, J., & Park, S. (2005, Nov 1). Working memory impairments in schizophrenia: a meta-analysis. *Journal of abnormal psychology*, *114*(4), 599–611. doi: 10.1037/0021-843X.114.4.599

Luck, J. S., & Vogel, K. E. (1997, Nov 20). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Lundqvist, M., Herman, P., & Miller, K. E. (2018, Aug 8). Working memory: Delay activity, yes! persistent activity? maybe not. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *38*(32), 7013–7019. doi: 10.1523/JNEUROSCI.2485-17.2018

Ma, J. W., Beck, M. J., Latham, E. P., & Pouget, A. (2006, Nov 22). Bayesian inference with probabilistic population codes. *Nature neuroscience*, *9*, 1432–1438. doi: 10.1038/nn1790

Ma, J. W., Husain, M., & Bays, M. P. (2014). Changing concepts of working memory. *Nature neuroscience*, *17*, 347–356.

Ma, J. W., Shen, S., Dziugaite, G., & van den Berg, R. (2015). Requiem for the max rule?

Vision research, *116*, 179–193.

Mackey, E. W., Winawer, J., & Curtis, E. C. (2017, Jun 19). Visual field map clusters in human frontoparietal cortex. *eLife*, *6*. doi: 10.7554/eLife.22974

Maloney, T. L., & Mamassian, P. (2009, Feb 5). Bayesian decision theory as a model of human visual perception: testing bayesian transfer. *Visual neuroscience*, *26*(1), 147–155. doi: 10.1017/S0952523808080905

Maloney, T. L., & Zhang, H. (2010, Nov 23). Decision-theoretic models of visual perception and action. *Vision research*, *50*(23), 2362–2374. doi: 10.1016/j.visres.2010.09.031

Mazyar, H., van den Berg, R., & Ma, J. W. (2012, Jun 8). Does precision decrease with set size? *Journal of vision*, *12*(6). doi: 10.1167/12.6.10

Miller, A. G. (1956, Mar 1). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological review*, *63*, 81–97. doi: 10.1037/h0043158

Moeller, S., Yacoub, E., Olman, A. C., Auerbach, E., Strupp, J., Harel, N., & U?urbil, K. (2010, May 1). Multiband multislice ge-epi at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fmri. *Magnetic resonance in medicine*, *63*(5), 1144–1153. doi: 10.1002/mrm.22361

Nassar, R. M., Helmers, C. J., & Frank, J. M. (2018, Jul 1). Chunking as a rational strategy for lossy data compression in visual working memory. *Psychological review*, *125*(4), 486–511. doi: 10.1037/rev0000101

Nobre, C. A., Coull, T. J., Maquet, P., Frith, D. C., Vandenberghe, R., & Mesulam, M. M. (2004, Apr 1). Orienting attention to locations in perceptual versus mental representations. *Journal of cognitive neuroscience*, *16*(3), 363–373. doi: 10.1162/089892904322926700

Oberauer, K. (2002, May 1). Access to information in working memory: exploring the focus of attention. *Journal of experimental psychology. Learning, memory, and cognition*, *28*(3), 411–421. doi: 10.1037//0278-7393.28.3.411

Oberauer, K., & Lin, H.-Y. (2017, Nov 21). An interference model of visual working memory. *Psychological review*.

Orhan, E. A., & Jacobs, A. R. (2014, Jul 2). Are performance limitations in visual short-term memory tasks due to capacity limitations or model mismatch? *arXiv*.

Park, C. D., Lautenschlager, G., Hedden, T., Davidson, S. N., Smith, D. A., & Smith,

K. P. (2002, Jun 1). Models of visuospatial and verbal memory across the adult life span. *Psychology and aging*, *17*(2), 299–320. doi: 10.1037//0882-7974.17.2.299

Phillips, A. W. (1974, Mar 1). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2). doi: 10.3758/BF03203943

Platt, L. M., & Glimcher, W. P. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*(6741), 233–238. doi: 10.1038/22268

Posner, I. M., Snyder, R. C., & Davidson, J. B. (1980, Jun 1). Attention and the detection of signals. *Journal of experimental psychology*, *109*(2), 160–174.

Postle, R. B. (2006, Apr 28). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23–38. doi: 10.1016/j.neuroscience.2005.06.005

Pratte, S. M., Park, E. Y., Rademaker, L. R., & Tong, F. (2017). Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working memory. *Journal of experimental psychology. Human perception and performance*, *43*(1), 6–17. doi: 10.1037/xhp0000302

Rademaker, L. R., Tredway, H. C., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision*, *12*, 21–21.

Rahmati, M., Saber, T. G., & Curtis, E. C. (2018, Feb 6). Population dynamics of early visual cortex during working memory. *Journal of cognitive neuroscience*, *30*(2), 219–233. doi: 10.1162/jocn_a_01196

Reinhart, G. R. M., Heitz, P. R., Purcell, A. B., Weigand, K. P., Schall, D. J., & Woodman, F. G. (2012). Homologous mechanisms of visuospatial working memory maintenance in macaque and human: properties and sources. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *32*(22), 7711–7722. doi: 10.1523/JNEUROSCI.0215-12.2012

Saber, T. G., Pestilli, F., & Curtis, E. C. (2015, Jan 7). Saccade planning evokes topographically specific activity in the dorsal and ventral streams. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *35*(1), 245–252. doi: 10.1523/JNEUROSCI.1687-14.2015

Sadaghiani, S., Hesselmann, G., Friston, J. K., & Kleinschmidt, A. (2010). The relation of ongoing brain activity, evoked neural responses, and cognition. *Front Sys Neurosci*, *4*(20). doi: 10.3389/fnsys.2010.00020

Salthouse, A. T., Babcock, L. R., & Shaw, J. R. (1991). Effects of adult age on structural

and operational capacities in working memory. *Psychology and aging*, *6*(1), 118–127. doi: 10.1037//0882-7974.6.1.118

Sapir, A., d'Avossa, G., McAvoy, M., Shulman, L. G., & Corbetta, M. (2005, Dec 6). Brain signals for spatial attention predict performance in a motion discrimination task. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(49), 17810–17815. doi: 10.1073/pnas.0504678102

Schneegans, S., & Bays, M. P. (2017, Apr 5). Neural architecture for feature binding in visual working memory. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *37*(14), 3913–3925. doi: 10.1523/JNEUROSCI.3493-16.2017

Schwarz, G. (Ed.). (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Serences, T. J., & Yantis, S. (2006, Jan 28). Selective visual attention and perceptual coherence. *Trends in cognitive sciences*, *10*(1), 38–45. doi: 10.1016/j.tics.2005.11.008

Serences, T. J., & Yantis, S. (2007, Feb 2). Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. *Cerebral cortex (New York, N.Y.: 1991)*, *17*(2), 284–293. doi: 10.1093/cercor/bhj146

Seung, S. H., & Sompolinsky, H. (1993, Nov 15). Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(22), 10749–10753. doi: 10.1073/pnas.90.22.10749

Sheremata, L. S., & Silver, A. M. (2015). Hemisphere-dependent attentional modulation of human parietal visual field representations. *J Neurosci.*, *35*(2).

Sims, R. C. (2015). The cost of misremembering: Inferring the loss function in visual working memory. *Journal of vision*, *15*(3). doi: 10.1167/15.3.2

Sims, R. C., Jacobs, A. R., & Knill, C. D. (2012, Oct 3). An ideal observer analysis of visual working memory. *Psychological review*, *119*(4), 807–830. doi: 10.1037/a0029856

Somers, C. D., Dale, M. A., Seiffert, E. A., & Tootell, B. R. (1999, Feb 16). Functional mri reveals spatially specific attentional modulation in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(4), 1663–1668. doi: 10.1073/pnas.96.4.1663

Sperling, G. (1960). The information available in brief visual presentations. *Psychological*

Monographs: General and Applied, *74*(11), 1–29.

Sprague, C. T., Ester, F. E., & Serences, T. J. (2014, Sep 22). Reconstructions of information in visual spatial working memory degrade with memory load. *Current biology: CB*, *24*(18), 2174–2180. doi: 10.1016/j.cub.2014.07.066

Sprague, C. T., Ester, F. E., & Serences, T. J. (2016, Aug 3). Restoring latent visual working memory representations in human cortex. *Neuron*, *91*(3), 694–707. doi: 10.1016/j.neuron.2016.07.006

Sprague, C. T., Itthipuripat, S., Vo, A. V., & Serences, T. J. (2018, Jun 1). Dissociable signatures of visual salience and behavioral relevance across attentional priority maps in human cortex. *Journal of neurophysiology*, *119*(6), 2153–2165. doi: 10.1152/jn.00059.2018

Sreenivasan, K. K., Curtis, E. C., & D'Esposito, M. (2014, Feb 14). Revisiting the role of persistent neural activity during working memory. *Trends in cognitive sciences*, *18*(2), 82–89. doi: 10.1016/j.tics.2013.12.001

Standing, L. (1973, May 1). Learning 10,000 pictures. *The Quarterly journal of experimental psychology*, *25*(2), 207–222. doi: 10.1080/14640747308400340

Suchow, W. J., Fougnie, D., & Alvarez, A. G. (2017). Looking inward and back: Real-time monitoring of visual working memories. *Journal of experimental psychology. Learning, memory, and cognition*, *43*(4), 660–668. doi: 10.1037/xlm0000320

Takikawa, Y., Kawagoe, R., Itoh, H., Nakahara, H., & Hikosaka, O. (2002). Modulation of saccadic eye movements by predicted reward outcome. *Experimental brain research*, *142*(2), 284–291. doi: 10.1007/s00221-001-0928-1

Thompson, G. K., & Bichot, P. N. (2005). A visual salience map in the primate frontal eye field. *Progress in brain research*, *147*, 251–262. doi: 10.1016/S0079-6123(04)47019-8

Todd, J. J., & Marois, R. (2004, Apr 15). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*(6984), 751–754. doi: 10.1038/nature02466

Treisman, A., & Sato, S. (1990, Aug 1). Conjunction search revisited. *Journal of experimental psychology. Human perception and performance*, *16*(3), 459–478. doi: 10.1037//0096-1523.16.3.459

Trommershäuser, J., Maloney, T. L., & Landy, S. M. (2003). Statistical decision theory and trade-offs in the control of motor response. *Spatial vision*, *16*(3-4), 255–275.

van Beers, J. R., Baraduc, P., & Wolpert, M. D. (2002, Aug 29). Role of uncertainty in sensorimotor control. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *357*(1424), 1137–1145. doi: 10.1098/rstb.2002.1101

van den Berg, R., Awh, E., & Ma, J. W. (2014, Jan 1). Factorial comparison of working memory models. *Psychological review*, *121*, 124–149. doi: 10.1037/a0035234

van den Berg, R., & Ma, J. W. (2018, Aug 7). A resource-rational theory of set size effects in human visual working memory. *eLife*, *7*. doi: 10.7554/eLife.34963

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, J. W. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *PNAS*, *109*, 8780–8785.

van den Berg, R., Yoo, H. A., & Ma, J. W. (2017). Fechner's law in metacognition: a quantitative model of visual working memory confidence. *Psychol Rev*, *124*(2), 197–214. doi: 10.1037/rev0000060

Victor, D. J., & Conte, M. M. (2004, Mar 1). Visual working memory for image statistics. *Vision research*, *44*(6), 541–556. doi: 10.1016/j.visres.2003.11.001

Vo, A. V., Sprague, C. T., & Serences, T. J. (2017, Mar 22). Spatial tuning shifts increase the discriminability and fidelity of population codes in visual cortex. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *37*(12), 3386–3401. doi: 10.1523/JNEUROSCI.3484-16.2017

Vogel, K. E., McCollough, W. A., & Machizawa, G. M. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. doi: 10.1038/nature04171

Wandell, A. B., Dumoulin, O. S., & Brewer, A. A. (2007, Oct 25). Visual field maps in human cortex. *Neuron*, *56*(2), 366–383. doi: 10.1016/j.neuron.2007.10.012

Wilken, P., & Ma, J. W. (2004). A detection theory account of change detection. *Journal of Vision*, *4*, 1120–1135.

Williams, W. D., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision research*, *24*(1), 55–62. doi: 10.1016/0042-6989(84)90144-5

Xu, J., Moeller, S., Auerbach, J. E., Strupp, J., Smith, M. S., Feinberg, A. D., ... U?urbil, K. (2013, Dec 27). Evaluation of slice accelerations using multiband echo planar imaging at 3 t. *NeuroImage*, *83*, 991–1001. doi: 10.1016/j.neuroimage.2013.07.055

Xu, Y. (2017, Oct 31). Reevaluating the sensory account of visual working memory storage. *Trends in cognitive sciences*, *21*(10), 794–815. doi: 10.1016/j.tics.2017.06.013

Xu, Y., & Chun, M. M. (2006, Mar 2). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*(7080), 91–95. doi: 10.1038/nature04262

Yoo, H. A., Klyszejko, Z., Curtis, E. C., & Ma, J. W. (2018, Nov 1). Strategic allocation of working memory resource. *Scientific reports*, *8*(1). doi: 10.1038/s41598-018-34282-1

Zhang, W., & Luck, J. S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.