12-2014

Change Detection in Rhesus Monkeys and Humans

Deepna T. Devkar

Deepna T. Devkar

CHANGE DETECTION IN RHESUS MONKEYS AND HUMANS

by

DEEPNA THAKKAR DEVKAR, M.S.

APPROVED:

_____
Anthony Wright, Ph.D., Supervisory Professor

_____
Wei Ji Ma, Ph.D.

_____
Valentin Dragoi, Ph.D.

_____
Anne Sereno, Ph.D.

_____
James Pomerantz, Ph.D.

APPROVED:

_____
Dean, The University of Texas
Health Science Center at Houston
Graduate School of Biomedical Sciences

CHANGE DETECTION IN RHESUS MONKEYS AND HUMANS

A DISSERTATION

Presented to the Faculty of The University of Texas Health Science Center at Houston and The University of Texas M. D. Anderson Cancer Center Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Deepna Thakkar Devkar, M.S.

Houston, Texas

October 2014

deepna.85@gmail.com

ACKNOWLEDGMENTS

My journey through graduate school has been an extraordinary one and I have many people to thank for their guidance, encouragement, and support along the way. First credit belongs to my advisor, Dr. Anthony Wright, for his invaluable supervision over the past three and a half years. His mentorship has been an exceptional balance of thoughtful advice, profound scientific insight, meticulous attention to detail, and freedom to discover my own successes and failures in research. I owe sincere gratitude to him for his significant contributions in my thesis work, his extraordinary training in my development as a scientist, and his enthusiastic participation in all of my endeavors.

I am deeply indebted to my co-advisor, Dr. Wei Ji Ma, who has been an incredible resource of knowledge from day one. His area of expertise, patient coaching with theoretical concepts and technical skills, critical suggestions, and frequent hours of discussion have enormously shaped the foundation of this dissertation. I have grown immensely under his tutelage over the years and across many miles, and am sincerely appreciative of his vested interest in me. His unparalleled teaching abilities, industrious nature, and selfless investment in his students are among his many qualities that will continue to be a source of inspiration in my academic pursuits.

I am truly fortunate to have an extremely helpful and supportive advisory committee. I am thankful to each of them for their insightful feedback and suggestions on my research projects and for investing their time to write stellar

recommendation letters for me on several occasions. I am especially thankful to Drs. Valentin Dragoi and Anne Sereno for being enthusiastic supporters in all of my milestones during graduate school, starting from my very first day as an interviewee. I have always found their doors to be open for any discussions.

To all the past members and affiliates of the lab, thank you for all the stimulating discussions and remote assistance over the years. I especially thank Caitlin Elmore and John Magnotti for their enduring friendship, many fun conversations over coffee breaks and lunches, and for always being just a message away from my instant rescue. To my friends and classmates, thank you for being my fun accomplices in our journey. My experience here would not have been half the fun without all our laughter at seemingly scary and daunting aspects of the academic process.

To my amazing parents to whom I dedicate this work, thank you for giving me the encouragement to run after my ambitions, and for realizing my potential before I ever did. I hope this work makes you proud. To my cousins (siblings-at-heart), Kavita and Ashka, and other close family members, thank you for vicariously living through my ups and downs and for providing the necessary emotional sustenance along the way. Finally, I am in continual debt to my husband, Dinesh, who has been my rock-solid support system over the years. His cheerful persona, sensible understanding, endless patience, and special ability to make me laugh at meaningless things have truly made this journey seem easier. His confidence in me fuels the excitement to pursue my new beginnings.

CHANGE DETECTION IN RHESUS MONKEYS AND HUMANS

Deepna Devkar, M.S.

Supervisory Professor: Anthony Wright, Ph.D.

Visual working memory (VWM) is the temporary retention of visual information and a key component of cognitive processing. The classical paradigm for studying VWM and its encoding limitations has been change detection. Early work focused on how many items could be stored in VWM, leading to the popular theory that humans could remember no more than 4±1 items. More recently, proposals have suggested that VWM is a noisy, continuous resource distributed across virtually all items in the visual field, resulting in diminished memory quality rather than limited quantity. This debate about the nature of VWM has predominantly been studied with humans. Nevertheless, nonhuman species could add a great deal to the debate by providing evidence related to evolutionary continuity (similarities and/or differences) and model systems for investigating the neural basis of VWM. To this end, in the first aim, we tested monkeys and humans in virtually identical change detection tasks, where the subjects identified which memory item had changed between two displays. In addition to the typical manipulation of the number of items to-be-remembered (2-5 oriented bars), we varied the change magnitude (degree of orientation change) – a critical manipulation for discriminating among leading

models of VWM encoding limitations. We found that in both species VWM performance was best accounted for by a model in which memory items are encoded in a noisy manner, where quality of memory is variable and on average decreases with increasing set size.

The second aim focused on the decision-making component of change detection, where observers use noisy sensory information to make a judgment about where the change occurred. We tested monkeys and humans in the same change detection task (Aim 1), but with ellipses that varied in their height-to-width ratio so that their reliability of communicating orientation discrimination could be manipulated. The high-reliability ellipses were long and narrow, and the low-reliability ellipses were short and wide. We compared models that differed with respect to how the observers incorporate knowledge of stimulus reliability during decision-making. We found that in both species performance was best accounted for by a Bayesian model in which observers take into account the uncertainty of sensory observations when making perceptual judgments, giving more weight to more reliable evidence.

The comparative results across these related primate species are suggestive of evolutionary continuity of basic VWM processing in primates generally. These findings provide a strong theoretical foundation for how VWM processes work and establish rhesus monkeys as a good animal model system for physiological investigations to elucidate the neural substrates of VWM processing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: GENERAL INTRODUCTION

In a brief instant, our visual system can be inundated with an overwhelming amount of information. The ability to store and process critical information efficiently from our rich, dynamic, and highly complex visual world is important to the survival of a species. Visual working memory (VWM) is the short-term retention and manipulation of visual information over a few seconds (Baddeley, 1992). It is a temporary buffer that allows the brain to compare information from the immediate past to the present and integrate changes in a visual scene (Phillips, 1974; Rensink, 2002).

A simple example of the importance of VWM in our everyday lives is when a car driver needs to make a quick decision about changing lanes: he/she must be able to detect changes in a traffic situation after looking in all directions (rear view mirror, side mirrors, front view of the road, etc.) and remember that information sufficiently so that it can be integrated to make safe and optimal decisions. Similarly, non-human animals constantly use VWM to detect changes in their visual scene to effectively navigate, forage, interact with conspecifics, and avoid predators.

Apart from its role in detecting changes in the visual scene, VWM also underpins the execution of many basic cognitive processes such as smooth visual perception across saccadic eye movements, target search, guidance of goal-directed reaching movements, and filtering of relevant information (Brouwer & Knill, 2007; Chun & Potter, 1995; Henderson, 2008; Irwin, 1991; Miller, Erickson, & Desimone, 1996; Rainer, Asaad, & Miller, 1998). Additionally, VWM is interlinked with visual attention, frontal executive control centers, and long-term

visual memory (Awh, Vogel, & Oh, 2006; Chun, 2011; Cowan, 2011; Fukuda & Vogel, 2011). Because vision is the dominant sense of many animals, including primates, VWM is thus fundamental to the cognition of such species.

In humans, performance on VWM tasks has been correlated with measures of higher cognitive abilities such as problem solving, learning, language comprehension, selective/executive attention, and general intelligence (Conway, Kane, & Engle, 2003; Cowan, 2001; Cowan et al., 2005; Kiyonaga & Egner, 2014). Given its importance in everyday cognitive functioning, it is not surprising that deficits in VWM have been associated with several cognitive and neuropsychiatric disorders such as spatial neglect, parietal and temporal lobe damage, Schizophrenia, Attention Deficit Hyperactivity Disorder, Autism, Alzheimer's disease, Post-traumatic Stress Disorder, and Depression (Alescio-Lautier et al., 2007; Berryhill & Olson, 2008; Christopher & MacDonald, 2005; Ezzyat & Olson, 2008; Farmer et al., 2000; Gold, Wilk, McMahon, Buchanan, & Luck, 2003; Kim, Liu, Glizer, Tannock, & Woltering, 2014; Pellicano, Gibson, Maybery, Durkin, & Badcock, 2005; Pisella, Berberovic, & Mattingley, 2004; Vasterling, Brailey, Constans, & Sutker, 1998). Despite decades of work relating VWM to psychiatric diseases and cognition generally, answers to many basic processes of VWM remain elusive. With so much left to be understood about the impairments associated with failures of VWM, a better understanding of the normal functioning of VWM mechanisms might provide a better foundation for treating these impairments and evaluating the efficacy of treatment.

Behavioral research aimed at understanding VWM has used delayed matching-to-sample, memory span, or *N*-back tasks which in many cases require remembering only a single memorandum at a given instant (Fuster & Alexander, 1971; Goldman-Rakic, 1995; Miller et al., 1996). Although these approaches have been influential in understanding some time limitations of VWM by testing only single-item VWM, they are not particularly relevant to natural visual scenes, where multiple items must be processed and integrated in a continuously changing stream of information (Funahashi, Bruce, & Goldman-Rakic, 1989; Fuster & Alexander, 1971; Goldman-Rakic, 1990; E. K. Miller et al., 1996). The basic processes by which multiple items are encoded and processed in visual working memory needs to be better understood for assessing many of the underlying processes, neural circuitry, and failures of VWM.

Over the past few decades, the leading task for investigating multiple-item VWM and the amount of information that can be maintained simultaneously in VWM has been change detection (Rensink, 2002). In a typical change detection task, an observer is presented with a sample array of two or more stimuli, which is followed by a brief delay (usually more than 80 ms to exceed the duration of attentional capture). The number of stimuli in the sample array (or the items that are to-be-remembered) will be referred to as set size. After the delay, a test array is presented with a changed item and the observer's task is to identify whether or where the change occurred between the two arrays.

Results from such human change detection studies have shown proportion correct to be very high for small set sizes (e.g., 2 - 4 items), but

becomes progressively less accurate with increasing set sizes beyond 3 to 4 items (Luck & Vogel, 1997; Pashler, 1988; Vogel, Woodman, & Luck, 2001). These results have led to the popular theory that VWM is capacity limited, where only a fixed number of items can be held in memory. This fixed-capacity theory was first suggested by George Miller; however at the time, the 'magical' number was thought to be 7± 2 items (Miller, 1956). The capacity was then estimated to be higher than it is now because of the human ability to "chunk" bits of information together to maximize capacity. This number has since been replaced with 4 ± 1 items (Cowan, 2001). This fixed-capacity theory has also been called the item-limit or slot theory because only a limited number of items are proposed to be stored in discrete "slots". According to this theory, items are encoded in memory in an all-or-none fashion such that remembered items are stored with high fidelity, and no information is retained about other items. This theory of a fixed capacity has dominated much of the thinking about working memory for about half-a-century and has formed the basis of many neural investigations of human VWM (Edward Awh, Barton, & Vogel, 2007; Luck & Vogel, 2013; Rouder et al., 2008; Todd & Marois, 2004; Vogel & Machizawa, 2004; Xu & Chun, 2006).

In the last decade, the item-limit model has been challenged in the human literature on several grounds. First of all, even though it has been argued that the capacity estimate is stable across several short-term memory modalities (including visual, verbal, and auditory) and across testing paradigms (Cowan, 2001, 2005), some studies have reported that this so-called 'magical number' actually does vary when information load and stimulus complexity are

manipulated (Alvarez & Cavanagh, 2004; Eng, Chen, & Jiang, 2005; Olson & Jiang, 2002). Second, the fixed-capacity theory proposes an absolute view of all-or-none storage that is highly questionable based on grounds of signal detection theory. Signal detection theory has dominated psychophysics for the past half century, showing that sensory observations are subject to noise, and detection performance is imperfect due to errors in separating the true signal from noise (Green & Swets, 1966; Macmillan & Creelman, 2005). Third, the notion that a stimulus can be encoded perfectly is at odds with the evidence that neural systems are inherently noisy (e.g., Faisal, Selen, & Wolpert, 2008).

An alternative theory that reconciles most of the problems associated with all-or-none fixed capacity is that memory is a continuous resource that can be allocated to many (if not all items) in the field of view (Ma, Husain, & Bays, 2014). At the inception of these resource models, Wilken & Ma (2004) proposed that stimuli are encoded in memory in a noisy fashion, with the level of noise per item increasing with set size. Memory precision (which is inversely related to noise) decreases with the number of objects in the visual scene. Thus, according to the resource view, performance decreases because of a reduction in the *quality* of memories, rather than a cap or limit on the *number* of items that can be stored (Bays & Husain, 2008; Bays, Catalao, & Husain, 2009; Keshvari, van den Berg, & Ma, 2013; van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004). Nevertheless, in an attempt to salvage the item-limit theory, variants on the item-limit theory have been proposed, including combining fixed capacity with resource (e.g., Zhang & Luck, 2008). Recent work in humans has attempted to

distinguish among the item-limit model, its more recent variants, and resource models (Keshvari et al., 2013; van den Berg, Awh, & Ma, 2014; van den Berg, Shin, et al., 2012; van den Berg & Ma, 2014).

Compared to this rich body of ongoing work in humans, very little is known about how visual information is encoded in non-human animals and whether their VWM system suffers from the same limitations as humans. Rhesus monkeys are an ideal species for such investigations because they have similar visual memory processing mechanisms as humans (Wright, 2007; Goldman-Rakic, 1990; Goldman-Rakic, 1995; Sands & Wright, 1980). Results from such studies might help to disambiguate some of the controversies surrounding visual memory processing mechanisms in primates generally. For example, if rhesus monkeys (or some other animal species) were to show qualitative similarity to humans in underlying mechanisms of VWM then this nonhuman animal could be used as a model system for invasive investigations of VWM such as electrophysiological recordings, lesions, genetic, and pharmacological manipulations.

Several recent studies have begun to investigate these questions in, rhesus monkeys, but the results and findings have been mixed (Buschman, Siegel, Roy, & Miller, 2011; Elmore et al., 2011; Heyselaar, Johnston, & Paré, 2011; Lara & Wallis, 2012).. For example, Elmore et al. (2011) noted that memory sensitivity (discriminability, *d*', used as a measure of precision) decreases as the number of memory items increases. Moreover, this decline in performance is well fit by a power law function. Similarly, Lara and Wallis (2012) reported that precision of memory representations, and subsequently,

performance accuracy decreases with increasing set size. These findings are consistent with the theory of a continuous-resource model, where memory resource can be flexibly allocated to multiple items. Buschman and colleagues simultaneously recorded from area V4 and prefrontal cortex in the rhesus macaque while the animal was performing a change localization task at varying set sizes. They found that at the neural level, information is distributed among multiple items in the visual scene in a graded fashion. However, this sharing of resource only occurred when items were displayed in the same visual hemifield. They concluded that the two hemifields seemed to have discrete, slot-like resources with independent capacities. To interpret neural data from such an experiment, it is essential to connect them to critical measures describing the animal's behavioral performance. However, which behavioral parameters are most relevant depends on which model describes behavior best. For example, if a resource model were to fit the behavioral results better than a fixed-capacity model, then the common practice of finding neural correlates of the item capacity would make little sense, and instead would point to a resource description of neural activity. Thus, it is essential to first determine which model best accounts for non-human primate behavior in order to specify the framework for neural investigations of the basis of VWM.

Unfortunately, psychophysical studies with monkeys are sparse and none have performed detailed model comparisons such as those in the human literature (Keshvari, van den Berg, & Ma, 2012; Keshvari et al., 2013; van den Berg, Shin, et al., 2012; but see Lara & Wallis, 2012 for limited model

comparisons). Ma and colleagues have suggested that in order to distinguish among models of VWM, it is important to measure change detection performance across a wide range of change magnitudes, in addition to the typical manipulation of set size. This approach has been used effectively with humans to distinguish among leading models of VWM in detailed model comparisons (Keshvari et al., 2012, 2013; van den Berg, Shin, et al., 2012). However, change detection studies with monkeys have typically used displays containing highly discriminable stimuli such as clip art images of everyday objects or colored shapes (Buschman et al., 2011; Elmore et al., 2011; Heyselaar et al., 2011; Lara & Wallis, 2012). Differences among such highly discriminable stimuli are difficult to measure (but see Elmore et al, 2011 for multidimensional scaling of stimuli). Nevertheless, measuring discriminability of such stimuli would not in itself provide a basis for distinguishing among current models of VWM. To make such distinctions, the degree of change discrimination needs to be parametrically varied to produce psychometric functions, where accuracy gradually rises from near chance performance (50% correct in many cases) to near maximum accuracy for very large change discriminations. To this end, in the first set of experiments to be presented, we used oriented line bars and systematically varied change magnitude along with set size. By using these manipulations and the psychometric functions generated by them, our purpose was to rigorously compare five leading models of VWM encoding in parallel with monkeys and humans in an identical change detection paradigm. The theory and

computational details of each of these models are described in detail in Chapter 2.

VWM processing consists of two components: an encoding stage, where internal representations of the observed stimulus are generated, and a decision-making stage, during which information from these noisy measurements is used to make a decision. Change detection tasks are designed to test both the encoding and decision components of VWM. The observer encodes information about the sample stimuli, compares the maintained memory of the sample stimuli with the test stimuli at the corresponding locations, makes a judgment about which test stimulus has changed, and then makes a response based on this decision. Memory of the stimuli are seldom perfectly precise, therefore this less-than-perfect precision translates (proportionately) to noisy internal representations of stimuli. This internal noise varies across stimuli and trials. For example, even when the same stimulus is presented repeatedly, the sensory responses that it evokes in the form of neural activity can vary largely from trial-to-trial (Faisal et al., 2008; Tolhurst, Movshon, & Dean, 1983). Thus, an observer has to make a judgment about sensory observations in the presence of uncertainty caused by both internal and external factors. In making such judgments, knowing the nature of memory precision would benefit the observer in making better decisions.

Signal detection theory suggests that observers use Bayesian inference to make decisions that will maximize his/her decision performance on a trial, given the noisy stimulus encoding. In the change detection task, for example, a

Bayesian observer uses this noisy information to compute a probability distribution of whether or where the change occurred. Based on this computation, the observer chooses the location with the highest probability of change. In recent decades, studies have shown that in many perceptual tasks, humans are Bayesian observers and take into account the uncertainty associated with the noisy encoding of stimuli (Knill & Pouget, 2004; Knill & Richards, 1996). This uncertainty about the stimulus and precision itself can vary from trial-to-trial. Thus, in order to optimize performance, the observer must interpret uncertain sensory information by taking into account memory precision on a trial-by-trial basis. This process is referred to as “probabilistic computation’’ (Ma, 2012).

Psychophysical evidence for these types of probabilistic computations have been reported across several paradigms, including change detection, cue combination, multisensory integration, object perception, and sensorimotor learning (Angelaki, Gu, & DeAngelis, 2009; Ernst & Banks, 2002; Kersten, Mamassian, & Yuille, 2004; Keshvari et al., 2012; Knill & Richards, 1996; Körding & Wolpert, 2004). In these paradigms, the encoding precision not only varies from trial to trial for the same stimulus but also varies across different stimuli and possibly other factors as well (e.g., location). The purpose of explicitly manipulating the nature of the stimuli themselves (in addition to their orientation) is to vary the reliability of the stimulus, such as the height-to-width aspect ratio of an ellipse or the contrast ratio of the stimuli. For example, shorter, wider ellipses would provide less reliable information about orientation and consequently less information about orientation changes than longer, narrower ellipses. A

Bayesian observer would give more weight to measurements with higher reliability and thus higher certainty. It is important to note that Bayesian inference does not always translate to optimal inference (Ma, 2012). Bayesian inference is based on a subjective computation over sensory observations, which are prone to incorrect assumptions. When an observer's Bayesian estimation is based on incorrect assumptions, he/she can be suboptimal. The question, then, is the degree to which observers optimally evaluate the reliability of the stimulus in making a task-relevant decision.

Aim 2 of my thesis focuses on the decision-making component of VWM in a task related to that of Aim 1. I tested monkeys and humans in the same decision task and compared three Bayesian models of decision-making that vary with respect to the assumption that the observer makes about memory precision, based on their evaluation of stimulus uncertainty. The theoretical explanations of these models and their mathematical derivations are described in Chapter 4.

Qualitative similarities in monkeys and humans for encoding (Aim 1) and decision-making (Aim 2) would suggest evolutionary continuity and provide a model system for exploring the neurobiology and neural circuitry of VWM. Differences could also be important. For example, monkeys might be similar to humans in stimulus processing (encoding) but different from humans in decision optimality. Such results might suggest a judgment difference despite processing similarity, and perhaps lead to a better understanding of how these fundamental cognitive processes are employed and how they might be improved.

CHAPTER 2: MODEL FORMALISM

As discussed in the previous chapter, VWM processing consists of two stages: an encoding stage, where internal representations of visually presented stimuli are generated in memory, and a decision stage, where the internal representations are used to make a decision. We conducted two experiments of change detection that were designed to tap into both of these stages of processing. In Aim 1, we tested leading models of VWM encoding. In this task, observers briefly viewed a sample array of *N* randomly oriented bars (henceforth called items) and, following a delay, a test array containing two randomly chosen items from the sample array, of which one had a different orientation than in the sample array. The magnitude of the associated orientation change could take one of nine values. Observers reported which item had changed orientation. In this chapter I describe how we mathematically formalized and tested the encoding and decision stages.

Encoding

We tested five leading models of encoding, which differ in the way that they conceptualize the precise nature of memory resource and how it is allocated across multiple items in the visual scene. The five models are: item-limit, equal precision, equal precision with a fixed capacity, variable precision, and variable precision with a fixed capacity. The theory and modeling of each of these models are described below. *N* here represents set size, or the number of items in the sample array and *K* represents capacity.

Item-limit (infinite-precision) model

In the item-limit model, observers cannot store more than *K* items. When $N \le K$, all items are stored. The probability of being correct is then $1-\varepsilon$, where ε accounts for lapses of attention and unintended responses. When $N > K$, *K* randomly selected items from the sample display are stored. When the test display appears, there are then four scenarios to consider:

- Both test items were stored. This happens with probability $\frac{K(K-1)}{N(N-1)}$. The probability of being correct is then $1-\varepsilon$.
- One test item was stored, the other was not. This happens with probability $2\frac{K(N-K)}{N(N-1)}$. The probability of being correct is then $1-\varepsilon$.
- Neither test item was stored. This happens with probability $\frac{(N-K)(N-K-1)}{N(N-1)}$ The observer then has to guess about which item changed, and probability correct is 0.5.

  Overall proportion correct is then

$$PC(N,K) = \frac{K(K-1)}{N(N-1)}(1-\varepsilon) + 2\frac{K(N-K)}{N(N-1)}(1-\varepsilon) + \frac{(N-K)(N-K-1)}{N(N-1)} \cdot 0.5$$
$$= 1-\varepsilon - \frac{(N-K)(N-K-1)}{N(N-1)}(0.5-\varepsilon)$$

Note that storing all N items (K=N) yields the same proportion correct, namely $1-\varepsilon$, as storing only $N-1$ items, since even if one test item was not stored, the trial can be answered correctly by using the other test item. As can be seen from the equation, in the item-limit model, proportion correct depends on set size but not on change magnitude.

Noise-based (finite-precision) models

We assume that both orientations in the test display, which we denote by φ_1 and φ_2, are known noiselessly to the observer, because they remain on the screen until the subject responds. We model the memories of the orientations in the sample display as noisy. Noise can stem from encoding (presentation time was limited) or maintenance of memories; we do not distinguish between these possibilities. We model the noisy memory of the i^{th} item in the sample display, denoted x_i (i=1,..,N), as following a Von Mises distribution (circular Gaussian distribution because our stimuli exist in circular space) centered at the true stimulus, θ_i, with concentration parameter κ_i:

$$p(x_i \mid \theta_i) = \frac{1}{2\pi I_0(\kappa_i)} e^{\kappa_i \cos(x_i - \theta_i)}. \qquad (1)$$

where I_0 is the modified Bessel functions of the first kind of order 0. We have postulated previously that the role of precision is played by the Fisher information in this memory representation, denoted J_i (Keshvari et al., 2013; van den Berg, Shin, et al., 2012). This quantity is related to the concentration parameter through

$$J_i = \kappa_i \frac{I_1(\kappa_i)}{I_0(\kappa_i)},$$

where I_1 is the modified Bessel functions of the first kind of order 1. The relationship between precision and concentration parameter is nearly the identity and none of our results would qualitatively change if one were to replace J_i by κ_i.

In the *equal-precision* (EP) model, the precision of each item is inversely related to set size through a power law:

$$J_i = \frac{J_{N=1}}{N^{\alpha}}.$$

where $J_{N=1}$ is the precision of a single item. The precision of all items in a display is equal.

In the *equal precision with fixed capacity* (EPF) model (also known as *slots-plus-resources*), no more than K items can be stored. Thus, the number of stored items is min(N,K). The precision of a stored item is inversely related to the number of stored items through a power law,

$$J_i = \frac{J_{N=1}}{\min(N,K)^{\alpha}}.$$

The precision associated with a non-stored item is zero. When $N \leq K$, the EPF model is equal to the EP model.

In the *variable-precision* (VP) model, precision exhibits fluctuations across both space and time. The precision of each item is drawn independently from a

gamma distribution with mean $\bar{J}$ and scale parameter τ. This mean is inversely related to set size through a power law:

$$\bar{J} = \frac{\bar{J}_{N=1}}{N^{\alpha}},$$

where $\bar{J}_{N=1}$ is the mean precision of a single item.

The *variable precision with fixed capacity* (VPF) model is equal to the VP model when $N \leq K$. The precision of a stored item is drawn independently from a gamma distribution with mean $\bar{J}$ and scale parameter τ. This mean is inversely related to the number of stored items through a power law:

$$\bar{J} = \frac{\bar{J}_{N=1}}{\min\left(N,K\right)^{\alpha}}$$

where $\bar{J}_{N=1}$ is the mean precision of a single item. The precision associated with a non-stored item is zero.

The models have 2, 2, 3, 3, and 4 free parameters, respectively.

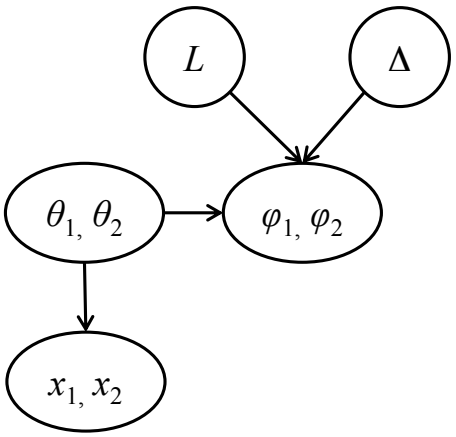
Decision-making

So far, we have described the encoding stage: how stimuli give rise to noisy memories. What is also needed in each of the noise-based models is a description of how the observer makes the two-alternative localization decision based on the noisy memories and the test display. We use an ideal (Bayesian-optimal) observer to describe this process. Bayesian-optimal inference refers to

the decision strategy that maximizes the observer's accuracy on a given trial based on the noisy measurements (Knill & Richards, 1996). The resulting decision rule is similar to the ideal-observer models of related *N*-alternative change localization and change detection tasks (Keshvari et al., 2012, 2013; van den Berg, Shin, et al., 2012), but differs in the details.

Step 1: Generative model

We begin by describing the decision process for the EP and VP models. The diagram shows the relevant variables: the location of the change, L (1 or 2), the magnitude of the change, Δ, the relevant sample orientations, θ_1 and θ_2 (all other sample items are irrelevant to the decision), their noisy memories, x_1, and x_2, and the two test orientations, φ_1 and φ_2.



Each variable has an associated probability distribution. Since both test locations are equally likely to contain the change, we have $p(L)=0.5$. Change magnitude Δ and each of the sample orientations have discrete distributions, but we

approximate them by uniform distributions, $p(\Delta) = \frac{1}{2\pi}$, and $p(\theta_1, \theta_2) = \left(\frac{1}{2\pi}\right)^2$.

The noisy memories x_1 and x_2 are distributed according to $p(x_1, x_2 \mid \theta_1, \theta_2) = p(x_1 \mid \theta_1) p(x_2 \mid \theta_2)$, where $p(x_i|\theta_i)$ is given by Eq. (1). Finally, the test orientations are $(\varphi_1, \varphi_2) = (\theta_1, \theta_2) + \Delta \mathbf{1}_L$, where 1_L is equal to (1,0) when L=1 and (0,1) when L=2.

Step 2: Inference

The observer infers L based on the noisy memories x_1 and x_2 and the test orientations φ_1 and φ_2. An ideal observer does this by computing the posterior distribution over L, $p(L|x_1,x_2,\varphi_1,\varphi_2)$. Since L is binary, all information about the posterior is contained in the log posterior ratio, which can be rewritten using Bayes' rule:

$$\log \frac{p(L=1 \mid x_1, x_2, \varphi_1, \varphi_2)}{p(L=2 \mid x_1, x_2, \varphi_1, \varphi_2)} = \log \frac{p(L=1)}{p(L=2)} + \log \frac{p(x_1, x_2, \varphi_1, \varphi_2 \mid L=1)}{p(x_1, x_2, \varphi_1, \varphi_2 \mid L=2)}$$
$$= \log \frac{p(x_1, x_2, \varphi_1, \varphi_2 \mid L=1)}{p(x_1, x_2, \varphi_1, \varphi_2 \mid L=2)},$$

since $p(L=1)=p(L=2)$. We evaluate the likelihood of L=1:

$$
\begin{aligned}
p(x_1, x_2, \varphi_1, \varphi_2 \mid L=1) &= \iiint p(x_1 \mid \theta_1)\, p(x_2 \mid \theta_2)\, p(\varphi_1, \varphi_2 \mid \theta_1, \theta_2, \Delta, L=1)\, p(\Delta)\, d\theta_1 d\theta_2 d\Delta \\
&= \iiint p(x_1 \mid \theta_1)\, p(x_2 \mid \theta_2)\, \delta(\varphi_1 - \theta_1 - \Delta)\, \delta(\varphi_2 - \theta_2) \frac{1}{2\pi} d\theta_1 d\theta_2 d\Delta \\
&= \frac{1}{2\pi} \int p(x_1 \mid \theta_1 = \varphi_1 - \Delta)\, p(x_2 \mid \theta_2 = \varphi_2)\, d\Delta \\
&= \frac{1}{2\pi} \frac{1}{2\pi I_0(\kappa_2)} e^{\kappa_2 \cos(x_2 - \varphi_2)} \int \frac{1}{2\pi I_0(\kappa_2)} e^{\kappa_2 \cos(x_1 - \varphi_1 + \Delta)} d\Delta \\
&= \frac{1}{2\pi} \frac{1}{2\pi I_0(\kappa_2)} e^{\kappa_2 \cos(x_2 - \varphi_2)}
\end{aligned}
$$

Similarly, the likelihood of *L*=2 is:

$$
p(x_1, x_2, \varphi_1, \varphi_2 \mid L=2) = \frac{1}{2\pi} \frac{1}{2\pi I_0(\kappa_1)} e^{\kappa_1 \cos(x_1 - \varphi_1)}.
$$

Combining, we find for the log posterior ratio

$$
\begin{aligned}
\log \frac{p(L=1 \mid x_1, x_2, \varphi_1, \varphi_2)}{p(L=2 \mid x_1, x_2, \varphi_1, \varphi_2)} &= \log \frac{\frac{1}{2\pi} \frac{1}{2\pi I_0(\kappa_2)} e^{\kappa_2 \cos(x_2 - \varphi_2)}}{\frac{1}{2\pi} \frac{1}{2\pi I_0(\kappa_1)} e^{\kappa_1 \cos(x_1 - \varphi_1)}} \\
&= \log \frac{I_0(\kappa_1)}{I_0(\kappa_2)} + \kappa_2 \cos(x_2 - \varphi_2) - \kappa_1 \cos(x_1 - \varphi_1)
\end{aligned}
$$

The ideal observer responds that the change occurred at location 1 when the log posterior ratio is positive:

$$
\log \frac{I_0(\kappa_1)}{I_0(\kappa_2)} + \kappa_2 \cos(x_2 - \varphi_2) - \kappa_1 \cos(x_1 - \varphi_1) > 0 \quad (2)
$$

This decision rule is valid for both the VP and EP models. In the VP model, precision per item is a random variable, and therefore κ_1 and κ_2 will

generally not be equal to each other. However, in the case of the EP model, we have $\kappa_1=\kappa_2$ and the inequality simplifies to

$$\cos\left(x_2-\varphi_2\right)>\cos\left(x_1-\varphi_1\right). \qquad (3)$$

This rule is intuitive: the observer reports that the change occurred at location 1 when the angular distance between the noisy memory at location 2 and the test orientation at location 2 is smaller than the corresponding distance at location 1. Then, there is more evidence that the change occurred at location 1. One can think of Eq. (2) as a precision-weighted version of Eq. (3).

The EPF model is very similar to the EP model, but with one difference when $N>K$. Then, a noisy measurement has a probability of not being stored. This is equivalent to setting the concentration parameter of the corresponding memory to 0. Thus, we can immediately obtain the decision rule from the EPF model by taking special cases of Eq. (2):

$$\begin{cases} \text{Report location 1 when....} \\ \cos\left(x_2-\varphi_2\right)>\cos\left(x_1-\varphi_1\right) \quad \text{if both items were stored;} \\ \kappa_1\cos\left(x_1-\varphi_1\right)<\log I_0\left(\kappa_1\right) \quad \text{if only item 1 was stored;} \\ \kappa_2\cos\left(x_2-\varphi_2\right)>\log I_0\left(\kappa_2\right) \quad \text{if only item 2 was stored;} \\ \quad \text{... and guess randomly when neither item was stored.} \end{cases} \qquad (4)$$

The VPF model is identical to the VP model when $N\leq K$. When $N>K$, just as in the EPF model above, a noisy measurement has a probability of not being stored (precision = 0). But, unlike the EPF model, the concentration parameters, κ_1 and κ_2 in the VPF model are drawn independently. With these modifications,

we can again take the special cases of Eq. (2) and obtain the decision rules for the VPF model:

$$\begin{cases} \text{Report location 1 when....} \\ \log\dfrac{I_0(\kappa_1)}{I_0(\kappa_2)} + \kappa_2\cos(x_2-\varphi_2) > \kappa_1\cos(x_1-\varphi_1) \text{ if both items were stored;} \\ \kappa_1\cos(x_1-\varphi_1) < \log I_0(\kappa_1) \quad \text{if only item 1 was stored;} \\ \kappa_2\cos(x_2-\varphi_2) > \log I_0(\kappa_2) \quad \text{if only item 2 was stored;} \\ \quad \text{... and guess randomly when neither item was stored.} \end{cases} \tag{5}$$

The second and third inequality in the EPF and VPF models may seem counterintuitive, since they only involve one memory. However, they make sense: even when the observer only has the memory corresponding to one of the test items, the discrepancy between the memory and the test is still informative about whether or not the change occurred in that one item.

Expected behavior

If we had access to the observer's noisy memories x_1 and x_2 on each trial, the model would predict their response exactly. Since we don't, the best we can do is to compute the *probability* of being correct for a given stimulus condition. Under the assumptions in our generative model, the stimulus condition is determined completely by set size N and change magnitude Δ, and the values of θ_1 and θ_2 are irrelevant. Thus, we are interested in the probability that the decision rule (Eq. (2) for VP, Eq. (3) for EP, Eq. (4) for EPF, and Eq. (5) for VPF) returns the correct location, when the memories x_1 and x_2 follow their respective

distributions given N and Δ. Without loss of generality, we compute proportion correct by taking $\theta_1=\theta_2=0$, and $L=1$, so that $\varphi_1=0$ and $\varphi_2=\Delta$.

For the EP model then,

$$\text{Proportion correct}\left(N,\Delta\right)=\Pr\left(\cos x_2>\cos\left(x_1-\Delta\right);x_1,x_2\sim\text{VM}\left(0,\kappa\right)\right)$$

where VM(μ,κ) denotes the Von Mises distribution with mean μ and concentration parameter κ.

For the EPF model, proportion correct is computed as a sum across the four possibilities for which items were stored (see Eq. (4)):

$$\begin{aligned}\text{Proportion correct}\left(N,\Delta\right)=&\frac{K\left(K-1\right)}{N\left(N-1\right)}\cdot\Pr\left(\cos x_2>\cos\left(x_1-\Delta\right);x_1,x_2\sim\text{VM}\left(0,\kappa\right)\ \right)\\&+\frac{K\left(N-K\right)}{N\left(N-1\right)}\cdot\Pr\left(\kappa\cos\left(x_1-\Delta\right)<\log I_0\left(\kappa\right);x_1\sim\text{VM}\left(0,\kappa\right)\right)\\&+\frac{K\left(N-K\right)}{N\left(N-1\right)}\cdot\Pr\left(\kappa\cos x_2>\log I_0\left(\kappa\right);x_2\sim\text{VM}\left(0,\kappa\right)\right)\\&+\frac{\left(N-K\right)\left(N-K-1\right)}{N\left(N-1\right)}\cdot0.5\end{aligned}$$

For the VP model,

$$\text{Proportion correct}\left(N,\Delta\right)=\Pr\left(\begin{array}{l}\log\frac{I_0\left(\kappa_1\right)}{I_0\left(\kappa_2\right)}+\kappa_2\cos x_2-\kappa_1\cos\left(x_1-\Delta\right)>0;\\x_1\sim\text{VM}\left(0,\kappa_1\right),x_2\sim\text{VM}\left(0,\kappa_2\right);J_i\sim\text{Gamma}\left(\bar{J},\tau\right)\end{array}\right)$$

For the VPF model, proportion correct is computed as a sum across the four possibilities for which items were stored (see Eq. (5)):

$$\text{Proportion correct}(N,\Delta) = \frac{K(K-1)}{N(N-1)} \cdot \Pr\left(\begin{array}{l} \log\frac{I_0(\kappa_1)}{I_0(\kappa_2)} + \kappa_2 \cos(x_2 - \varphi_2) > \kappa_1 \cos(x_1 - \varphi_1); \\ x_1 \sim \text{VM}(0,\kappa_1), x_2 \sim \text{VM}(0,\kappa_2); \\ J_i \sim \text{Gamma}(\bar{J},\tau) \end{array} \right)$$
$$+ \frac{K(N-K)}{N(N-1)} \cdot \Pr\left(\begin{array}{l} \kappa_1 \cos(x_1 - \Delta) < \log I_0(\kappa_1); x_1 \sim \text{VM}(0,\kappa_1); \\ J_i \sim \text{Gamma}(\bar{J},\tau) \end{array} \right)$$
$$+ \frac{K(N-K)}{N(N-1)} \cdot \Pr\left(\begin{array}{l} \kappa_2 \cos x_2 > \log I_0(\kappa_2); x_2 \sim \text{VM}(0,\kappa_2); \\ J_i \sim \text{Gamma}(\bar{J},\tau) \end{array} \right)$$
$$+ \frac{(N-K)(N-K-1)}{N(N-1)} \cdot 0.5$$

Each of these proportions correct was determined through Monte Carlo simulation, i.e. through a large number (10,000) of random draws of x_1 and x_2 (and of J_1 and J_2 as well in the case of the VP and VPF models). For each draw, we evaluated the decision rule, and then computed across all draws the proportion of correct responses.

Finally, for each model, we discretized parameter space finely and calculated a look-up table in which each entry gave the predicted probability of a correct response at one (N,Δ) combination for one parameter combination. Once we derived proportion correct for each model, the next step was to fit the models to the subjects' data. The specific methods for this step are described in the next chapter.

CHAPTER 3: CHANGE DETECTION TESTING IN RHESUS MONKEYS AND HUMANS: ENCODING

Introduction

As described in Chapter 2, we tested the five leading models of VWM encoding n parallel with monkeys and humans: 1) item-limit 2) equal-precision 3) equal-precision+fixed capacity 4) variable-precision and 5) variable-precision+fixed capacity (Figure 3.1).

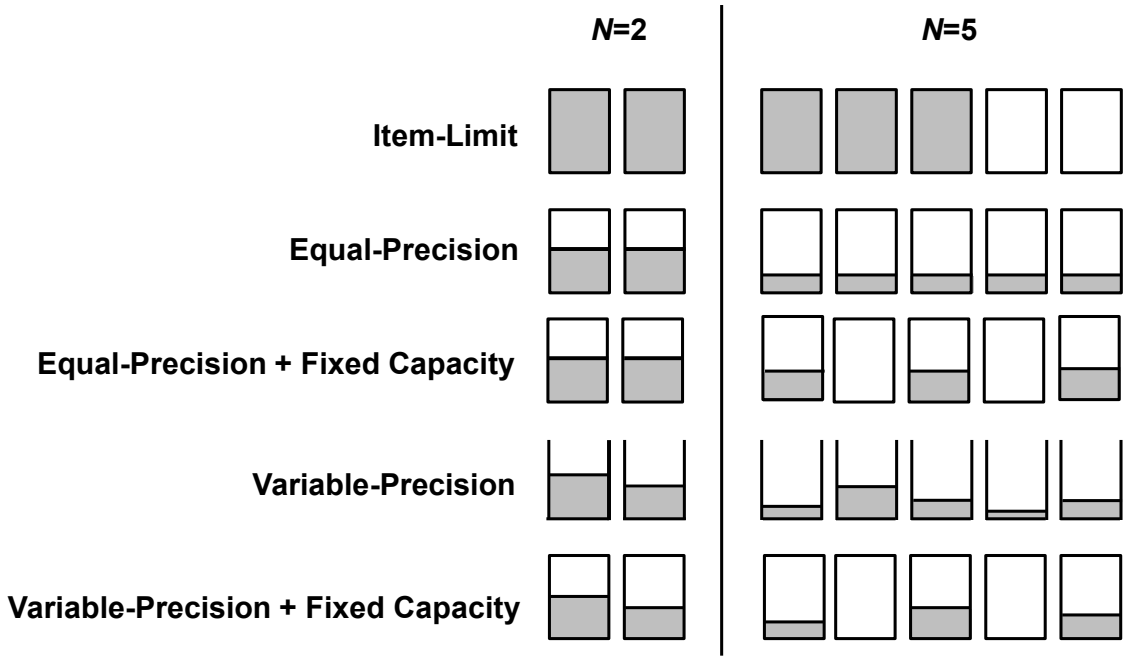


Figure 3.1 Schematic representation of resource allocation in five leading models of VWM. Each box represents an item to-be-remembered and the height of the fill represents the amount of resource allocated to that item. Set size is 2 (left) or 5 (right), with a hypothetical capacity limit of 3 for the Item-Limit, Equal-Precision + Fixed Capacity, and Variable-Precision + Fixed Capacity models.

According to the *item-limit* (IL) *model*, a fixed number of items (the capacity) are kept in memory with infinite precision, while remaining items are absent from memory (Cowan, 2001; S J Luck & Vogel, 1997; Pashler, 1988). The *equal-precision* (EP) *model* postulates that all items are remembered with equal memory precision, but the precision per item decreases with increasing set size (Palmer, 1990; Shaw, 1980). Decreasing precision is associated with increasing noise; that is, at a larger set size, each item is remembered in a noisier fashion (with lower precision). The *equal-precision + fixed capacity* (EPF) *model* combines elements of the item-limit and equal-precision models such that only a fixed number of items can be remembered, but each item in memory has finite precision (Zhang & Luck, 2008). When set size is smaller than the capacity, the model allows for precision to depend on set size. The *variable-precision* (VP) *model* is like the equal-precision model in that all items are remembered with finite precision but, by contrast, precision can vary from item to item and trial to trial (Fougnie, Suchow, & Alvarez, 2012; Keshvari et al., 2013; van den Berg, Shin, et al., 2012). The *variable-precision + fixed capacity* (VPF) *model* combines elements of the item-limit and variable-precision models such that only a fixed number of items can be remembered, but precision varies across items (van den Berg et al., 2014). All four finite-precision models (EP, VP, and to a lesser extent, EPF and VPF) attribute change-detection errors to the difficulty of separating the signal from noise. For these four models, we used Bayesian inference to model the decision stage; on each trial, the observer reports the location that has the

highest probability of containing the changed item (see Chapter 2. The IL, EP, EPF, VP, and VPF models have 2, 2, 3, 3 and 4 free parameters, respectively.

Three rhesus monkeys were tested for 11,520 trials each and ten humans were tested for 1152 trials each on the same visual change detection task (Figure 3.2). Subjects viewed a brief sample array of 2, 3, 4, or 5 randomly oriented bars (henceforth called items) and, following a delay, a test array containing two randomly chosen items from the sample array, of which one had a different orientation than in the sample array. The magnitude of the associated orientation change could take one of nine values. Subjects identified which item had changed orientation by touching it, and received trial-to-trial feedback.

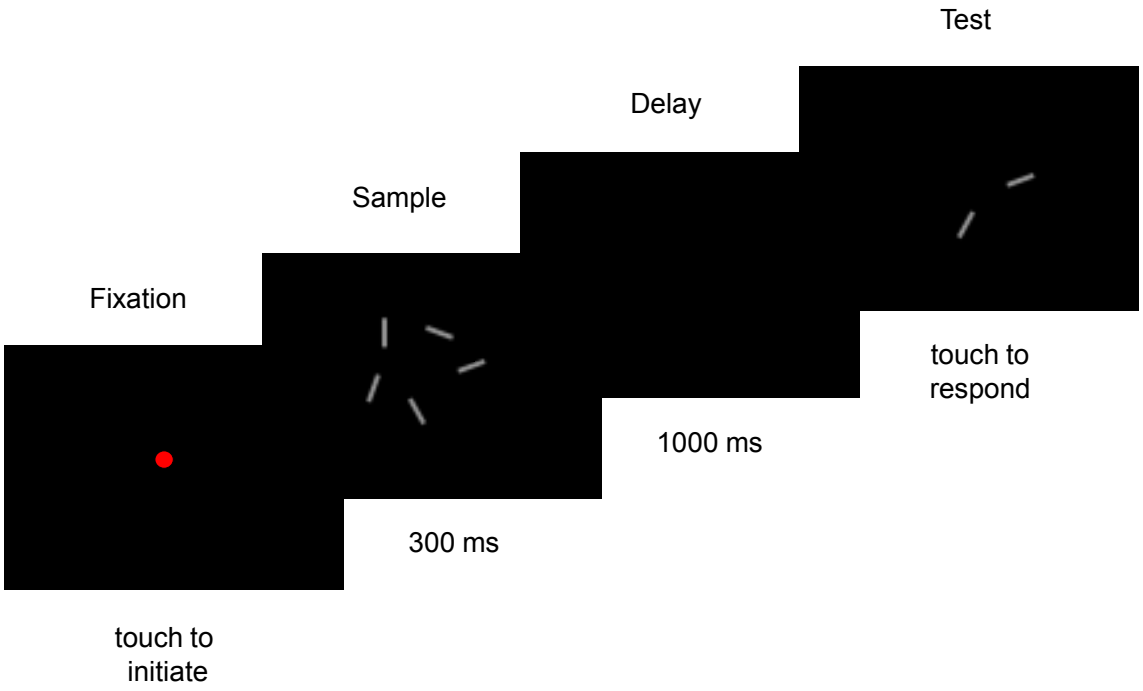


Figure 3.2 Trial Procedure. Subjects (monkeys and humans) were asked to report which item changed in its orientation between sample and test displays.

Methods

Monkeys

Subjects

Three adult male rhesus monkeys (*Macaca mulatta*; weights: M1 = 16.5 kg, M2 = 14.5 kg, M3 = 13.51 kg; ages: M1 = 17.5, M2 = 16.5, and M3 = 12.5 years) were tested in a change detection experiment for five days each week. Food and water were regulated prior to experimental sessions. After completing daily testing, animals were returned to their caging room, where they were housed individually and received primate chow and water to maintain their normal body weight. On days that the monkeys were not tested, they were given supplemental fruits and vegetables for enrichment in addition to the daily diet. All animal procedures were performed in accordance with the National Institutes of Health guidelines, approved by the Institutional Review Board at University of Texas Health Science Center at Houston, and supervised by the Institutional Animal Care and Use Committee.

Apparatus

During experimental sessions, the monkeys were placed unrestrained in a custom-made aluminum experimental chamber (47.5 cm wide x 53.1 cm deep x 66.3 cm high). An infrared touchscreen detected touch responses to a 17" computer monitor. The touch responses were guided using a Plexiglas template with 6 cutouts (diameter of each circle cutout = 2.75 cm) that were arranged on an imaginary circle of 9.0 cm diameter, matching the six possible locations of the

stimuli, and a cutout in the center (diameter = 2.5 cm) for touches to a fixation point. Using a computer-controlled relay interface (Model P10-12; Metrabyte, Taunton, MA), correct responses were rewarded with either a banana pellet or cherry Koolaid. The relay interface controlled the illumination of the chamber using a 25 W green light bulb located outside of the chamber. The offset of the green light illuminating the chamber through a small gap between the touchscreen and the monitor marked the start of the next trial. Throughout testing, the monkeys were monitored with a video camera outside the chamber and focused through a small glass covered port on the right side of the chamber. Experimental sessions were designed, operated, and recorded using a custom program written in Microsoft Visual Basic 6.0.

Stimuli

Stimuli consisted of 1.8 cm x 0.4 cm white bars displayed on a black background. Based on the average distance of the monkey from the screen (approximately 35 cm), the stimuli subtended a visual angle of 2.9° x 0.65°. Stimuli were presented in six possible locations on the screen, arranged on an imaginary circle (see Apparatus).

Trial procedure

Each trial began with a red fixation point in the center of the screen as shown in Figure 2.1. The monkeys had to make a one-touch response to the fixation point, which initiated the presentation of a sample display. This display contained two or more items (see below), and had a duration that differed

between monkeys and between training and testing (see below). After a delay of 1000 ms, the test display was presented, which always consisted of two items, placed at the same locations as two items from the sample display. One test item had the same orientation as the corresponding item in the sample display, and the other test item had a different orientation. The monkeys' task was to identify which item had changed, and to touch that item. The test display remained on the screen until response. Correct responses were rewarded. An intertrial interval of 3000 ms followed the choice response, during which a green light illuminated the chamber and the screen was dark.

Training

Two of the monkeys that participated in this study (M2 and M3) had been previously trained in a change detection task using clip art images and colored squares (Elmore et al., 2011). For these two monkeys, we intermixed trials of oriented bars (new stimuli) with trials of colored squares for initial task acquisition. Once the monkeys' performance on these orientation trials was similar to their baseline color trial performance, we began training them with only orientation trials. Since M1 had not been previously trained on this task, we directly trained him with oriented bars. All three monkeys were first trained at set sizes 2 and 3, change magnitudes of 22.5°, 45°, 67.5°, and 90°, and a sample viewing time of 1000 ms. Once overall accuracy reached approximately 70%, set sizes 4 and 5 and finer change magnitudes (10° to 90° in 10° increments) were gradually introduced. Finally, we gradually reduced sample viewing times while maintaining approximately 70% accuracy on set size 2 trials. For M1 and M3, this

led to a viewing time of 300 ms, and for M2 to a viewing time of 600 ms. Total training lasted approximately 8 months.

Testing

The sample display was shown for 300 ms for M1 and M3, and 600 ms for M2. Set size was 2, 3, 4, or 5. Set sizes were pseudorandomized within each 192-trial block (48 trials per set size). The orientation of each sample item, θ, was drawn independently from a uniform distribution over 18 possible orientations (−90°, −80°,…, −10°, 10°, 20°,…, 80°). The orientation of the changed item in the test display was drawn from the same distribution, except for orientation of the other sample items on this trial, so that the changed orientation would not be confused with other stimuli that did not change. Testing consisted of 60 sessions, with 192-trial blocks per session, for a total of 11,520 trials per monkey.

Humans

Subjects

Ten human subjects (8 females) aged 21-33 years (mean age = 27.1 years) participated. Each subject visited the lab for two 1.5-hour sessions and was compensated $10 per session. Study procedures were approved by the University of Texas Health Science Center at Houston Institutional Review Board.

Apparatus and stimuli

Subjects were seated in a chair in a small room equipped with a computer. At the beginning of the experiment, the distance between the chair and the screen was adjusted so that the stimuli and display would subtend approximately the same visual angles as for the monkeys. Subjects were asked to maintain approximately the same distance. The monitor and touchscreen were identical to those used for monkeys. Two 25 W light bulbs were mounted on the wall behind the subjects to provide feedback. Stimuli were identical to those used for monkeys.

Trial Procedure

The trial procedure was identical to that for the monkeys, except for the feedback. Feedback consisted of a green room light (75 W) that illuminated the testing room for 1 s and was accompanied by a tone following correct responses, or a red light that illuminated testing room for 1 s following incorrect responses.

Training and Testing

Each subject completed two testing sessions, each consisting of three 192-trial blocks, for a total of 1152 trials per subject. Subjects were given a 10-minute break time in between blocks. Each subject completed 8 practice trials at the beginning of the first session.

Model fitting

Denoting all parameters of a model by a vector **t**, the log likelihood of **t** (the parameter log likelihood) is

$$LL(\mathbf{t}) = \log p(\text{data} \mid \text{model}, \mathbf{t}) = \log \prod_{i=1}^{n_{\text{trials}}} p(\text{correctness}_i \mid N_i, \Delta_i, \mathbf{t})$$

where the product is over trials (from 1 to n_{trials}) and correctness$_i$ is 1 if the subject was correct on the i^{th} trial and 0 if not. We can rewrite this as

$$\begin{aligned} LL(\mathbf{t}) &= \sum_{i=1}^{n_{\text{trials}}} \log p(\text{correctness}_i \mid N_i, \Delta_i, \mathbf{t}) \\ &= \sum_{N} \sum_{\Delta} \begin{bmatrix} n(N, \Delta, \text{correct}) \cdot \log p(\text{correct} \mid N, \Delta, \mathbf{t}) + \\ + n(N, \Delta, \text{incorrect}) \cdot \log(1 - p(\text{correct} \mid N, \Delta, \mathbf{t})) \end{bmatrix}, \end{aligned} \quad (1)$$

where trials are grouped by set size N, change magnitude Δ, and by whether the observer was correct or incorrect, and n (N,Δ,correct) is the number of trials with a particular N, Δ, and correctness.

For each subject data set, we used Eq. (1) and the precomputed look-up table of model predictions mentioned above to find the log likelihood of each parameter combination. The parameter combination on this grid that maximized the log likelihood gives the estimates of the parameters. The model predictions corresponding to that parameter combination were then used to compute the model fits to the psychometric curves.

Model comparison

To compare models, we used four metrics: the Akaike Information Criterion (Akaike, 1974), the Akaike Information Corrected Criterion (Burnham, 2002; Hurvich & Tsai, 1989), the Bayesian Information Criterion (Schwarz, 1978), and

the log marginal likelihood (MacKay, 2003). All four measures penalize models for having more free parameters, but the penalties differ.

Akaike Information Criterion (AIC)

AIC rewards a model's good fit but penalizes free parameters. It is defined as

$$\mathrm{AIC} = -2LL_{\max} + 2k,$$

where $LL_{\max}$ is the maximum of the parameter log likelihood $LL(\mathbf{t})$ and k is the number of free parameters in the model. In this thesis, the following multiple of AIC are reported:

$$\mathrm{AIC}^* \equiv -\frac{1}{2}\mathrm{AIC} = LL_{\max} - k,$$

so that the leading term is the maximum log likelihood.

Akaike Information Corrected Criterion (AICc)

AICc is a corrected version of AIC, designed for data sets with few trials:

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2k(k+1)}{n_{\mathrm{trials}} - k - 1},$$

where n_{trials} denotes the number of trials. We report a modified AICc value, defined as

$$\mathrm{AICc}^* \equiv -\frac{1}{2}\mathrm{AICc} = LL_{\max} - k - \frac{k(k+1)}{n_{\mathrm{trials}} - k - 1}.$$

Bayesian Information Criterion (BIC)

BIC is similar to AIC in that it is also based on the maximum likelihood, but it has a larger penalty term for the number of free parameters.

$$\mathrm{BIC} = -2LL_{\max} + k \log n_{\mathrm{trials}}.$$

We report the modified BIC,

$$\mathrm{BIC}^{*} \equiv -\frac{1}{2}\mathrm{BIC} = LL_{\max} - \frac{k}{2}\log n_{\mathrm{trials}}.$$

Log Marginal Likelihood (LML)

Bayesian model comparison consists of calculating the log likelihood of a model *m* given the data, *LL*(model) = log *p*(data|model). Unlike the previous metrics, this is not solely based on the maximum likelihood. Instead, it involves integrating over the parameters; this is also called marginalizing over the parameters, which is why *LL*(model) is also called the log marginal likelihood:

$$\begin{aligned} LL(\mathrm{model}) &= \log p(\mathrm{data} \mid \mathrm{model}) \\ &= \log \int p(\mathrm{data} \mid \mathrm{model}, \mathbf{t})\, p(\mathbf{t} \mid \mathrm{model})\, d\mathbf{t} \\ &= \log \int e^{LL(\mathbf{t})}\, p(\mathbf{t} \mid \mathrm{model})\, d\mathbf{t} \end{aligned}$$

For the parameter prior $p(\mathbf{t}|\mathrm{model})$, we chose a discrete uniform distribution on the same grid as used for parameter estimation. We denote the size of the range of the j^{th} parameter by R_j, and its grid spacing by δt_j. Numerical values are specified in Table S1. We also rewrite slightly so as to avoid highly negative numbers in the exponent (those cause numerical underflow). Then we find

$$LL(\text{model}) = LL_{\max} + \sum_{j=1}^{k} \log \frac{\delta t_j}{R_j} + \log \sum_{\mathbf{t} \text{ on grid}} e^{LL(\mathbf{t}) - LL_{\max}} .$$

The difference of the log marginal likelihood between two models is also called the log Bayes factor of those two models.

Bootstrapping

Since we had only three monkey subjects, we used bootstrapping (Efron, 1993) for each monkey separately to estimate the standard errors on all summary statistics. The original data set for each monkey consisted of 11,520 trials. We sampled the same number of trials (a combination of set size, change magnitude, and correctness) with replacement from this dataset, to create a new dataset. We repeated sampling 100 times to create 100 bootstrapped data sets for each monkey. From each individual bootstrapped data set, we estimated the parameters, computed psychometric curves, calculated R^2, and computed AIC, AICc, BIC, and LML, and computed means for each by averaging across all bootstrapped data sets from the same monkey, with standard deviations serving as estimates of the standard errors of the means.

Results and Discussion

For both species, proportion correct decreased monotonically as a function of set size, with humans being substantially more accurate than monkeys (Figure 3.3).

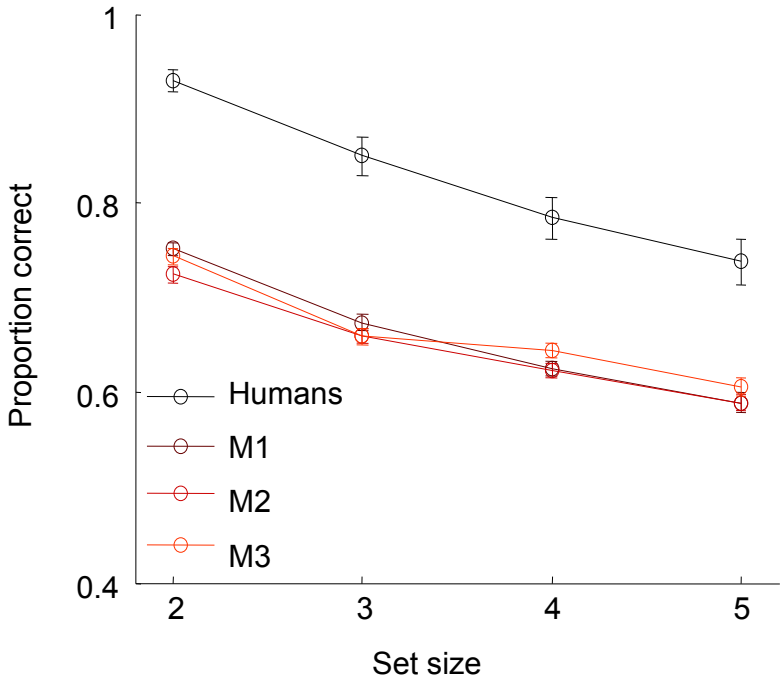


Figure 3.3 Proportion correct as a function of set size for humans and three monkeys (M1, M2, and M3).

A more detailed representation of the data is provided by proportion correct as a function of change magnitude for each of the four set sizes (Figure 3.4). We found large effects of both set size and change magnitude on VWM performance in both species (humans: two-way repeated-measures ANOVA; set size: $F(3,27) = 64.05$, $p < 0.001$, change magnitude: $F(8,72) = 80.36$, $p < 0.001$). The dependence on set size replicates findings found in many prior studies, both with humans (Keshvari et al., 2013; S J Luck & Vogel, 1997; van den Berg, Shin, et al., 2012; Wilken & Ma, 2004) and with monkeys (Buschman et al., 2011; Elmore et al., 2011; Heyselaar et al., 2011; Lara & Wallis, 2012). While most studies have ignored the variable of change magnitude, a few recent studies have systematically measured the effect of change magnitude and found effects similar to those shown in Figure 2.4 (Keshvari et al., 2012, 2013; Lara & Wallis, 2012; van den Berg, Shin, et al., 2012). Here, we exploit the statistical strength afforded by the joint dependencies of proportion correct on set size and change magnitude to compare models of VWM.

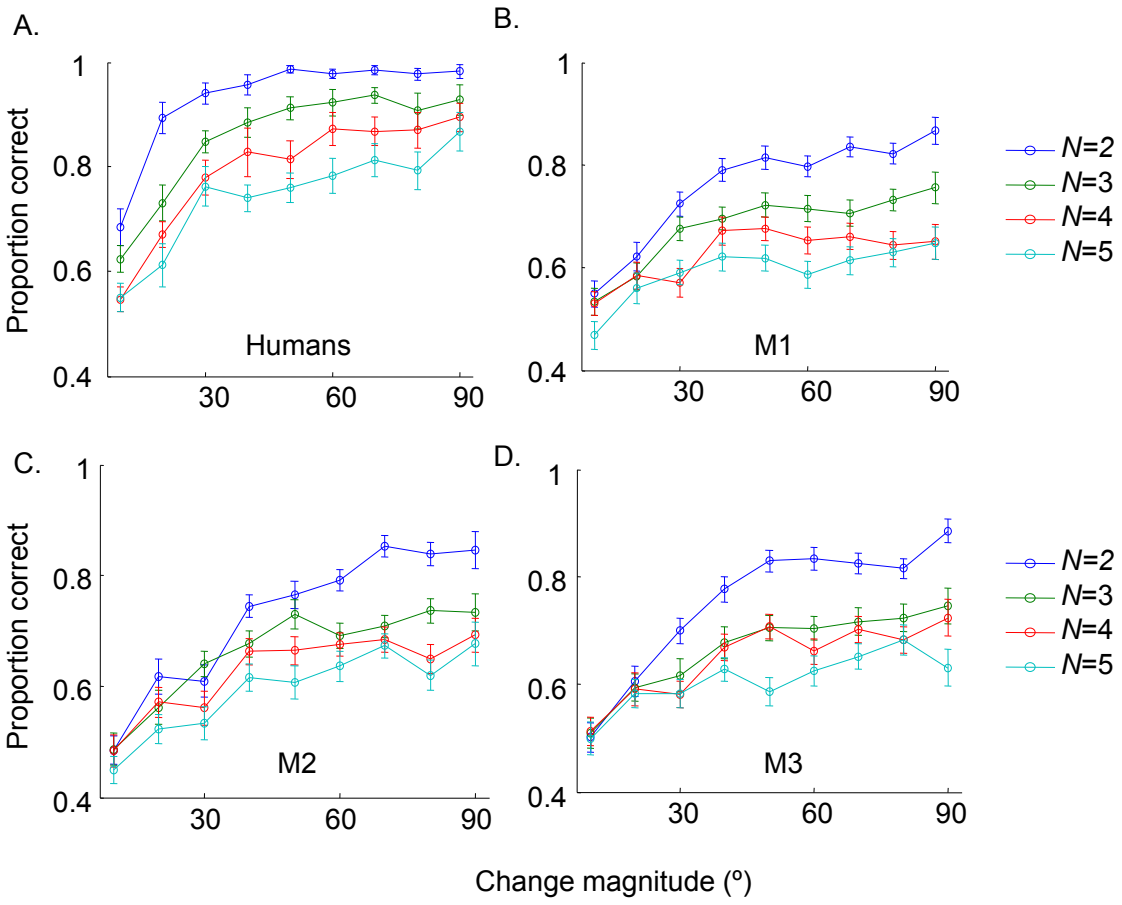


Figure 3.4 (A) Proportion correct across set size (*N*) and change magnitude (º) for humans; mean ± s.e.m. across ten subjects (B – D) Same for M1, M2, and M3 respectively; mean ± s.e.m. across bootstrapped datasets.

In spite of the large performance differences between both species, it is possible that the underlying VWM mechanisms are the same. To test this possibility, we compared the five leading models of VWM limitations for each individual monkey and human. We used maximum-likelihood estimation to fit the parameters in each model for each individual human subject as well as for each data set sampled using bootstrapping from an individual monkey's data. We found that models could not be strongly distinguished based on set size only (Figures 3.5-3.8).

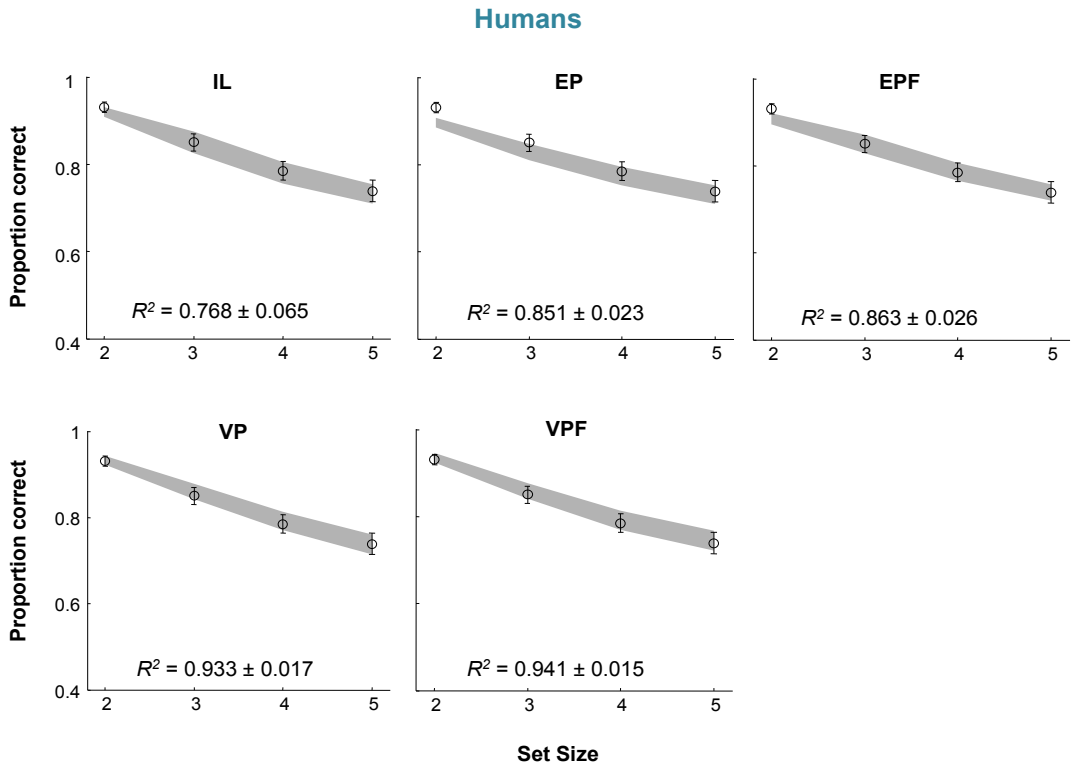


Figure 3.5. Proportion correct as a function of set size for humans. Circles and error bars: behavior; shaded areas: model fits. IL: Item-Limit, EP: Equal-Precision, EPF: Equal-Precision + Fixed Capacity, VP: Variable-Precision, VPF: Variable-Precision + Fixed Capacity.

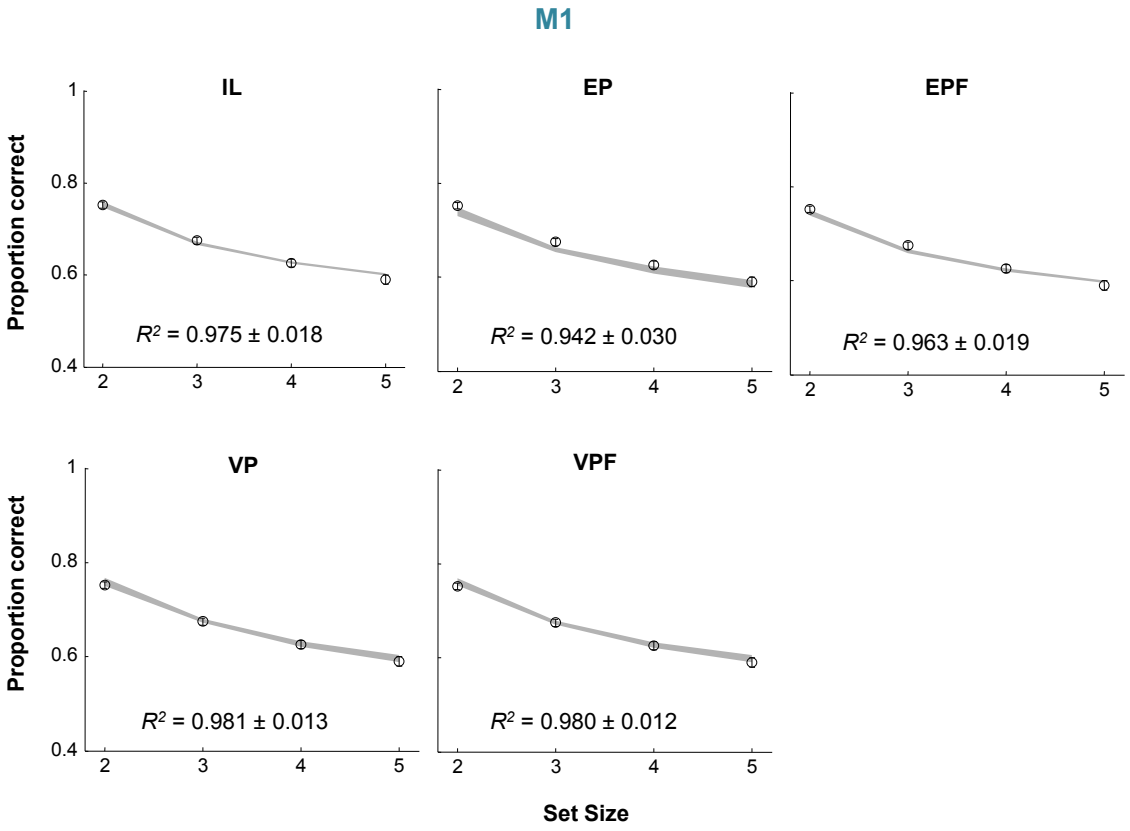


Figure 3.6. Proportion correct as a function of set size for M1. Circles and error bars: behavior; shaded areas: model fits. IL: Item-Limit, EP: Equal-Precision, EPF: Equal-Precision + Fixed Capacity, VP: Variable-Precision, VPF: Variable-Precision + Fixed Capacity.

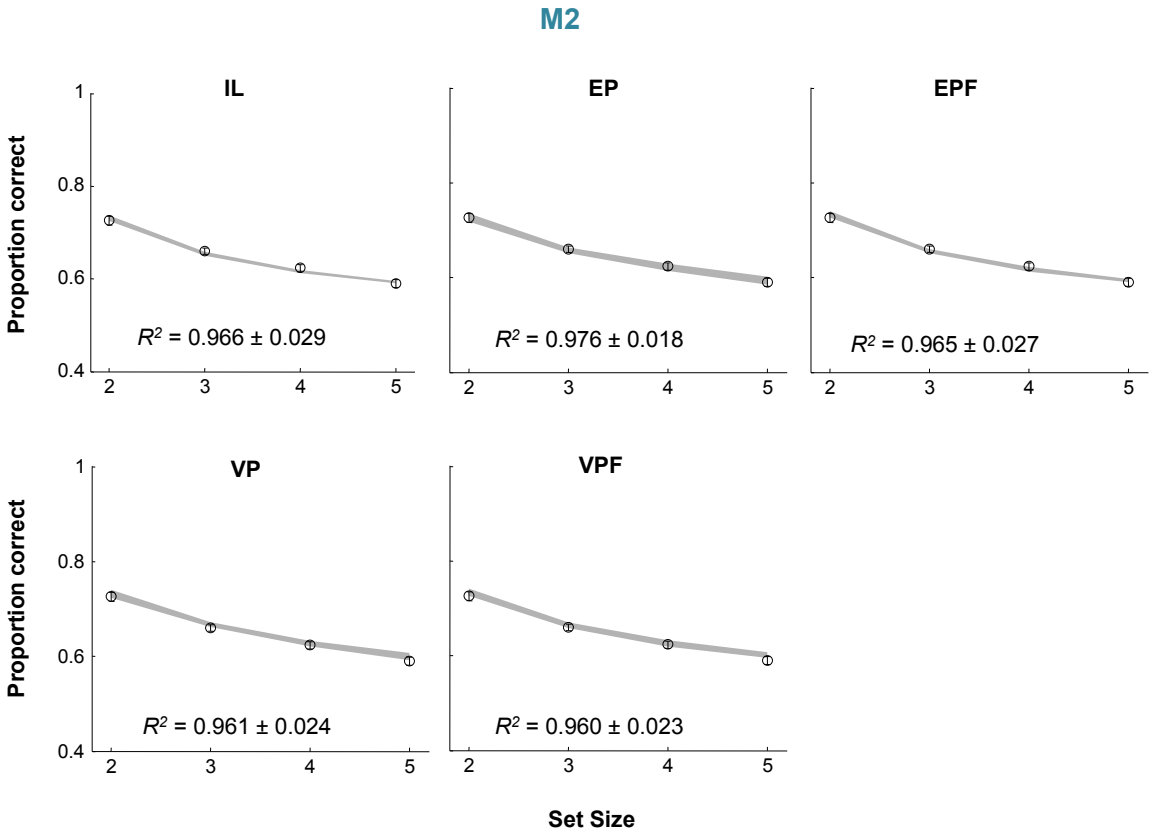


Figure3.7. Proportion correct as a function of set size for M2. Circles and error bars: behavior; shaded areas: model fits. IL: Item-Limit, EP: Equal-Precision, EPF: Equal-Precision + Fixed Capacity, VP: Variable-Precision, VPF: Variable-Precision + Fixed Capacity.

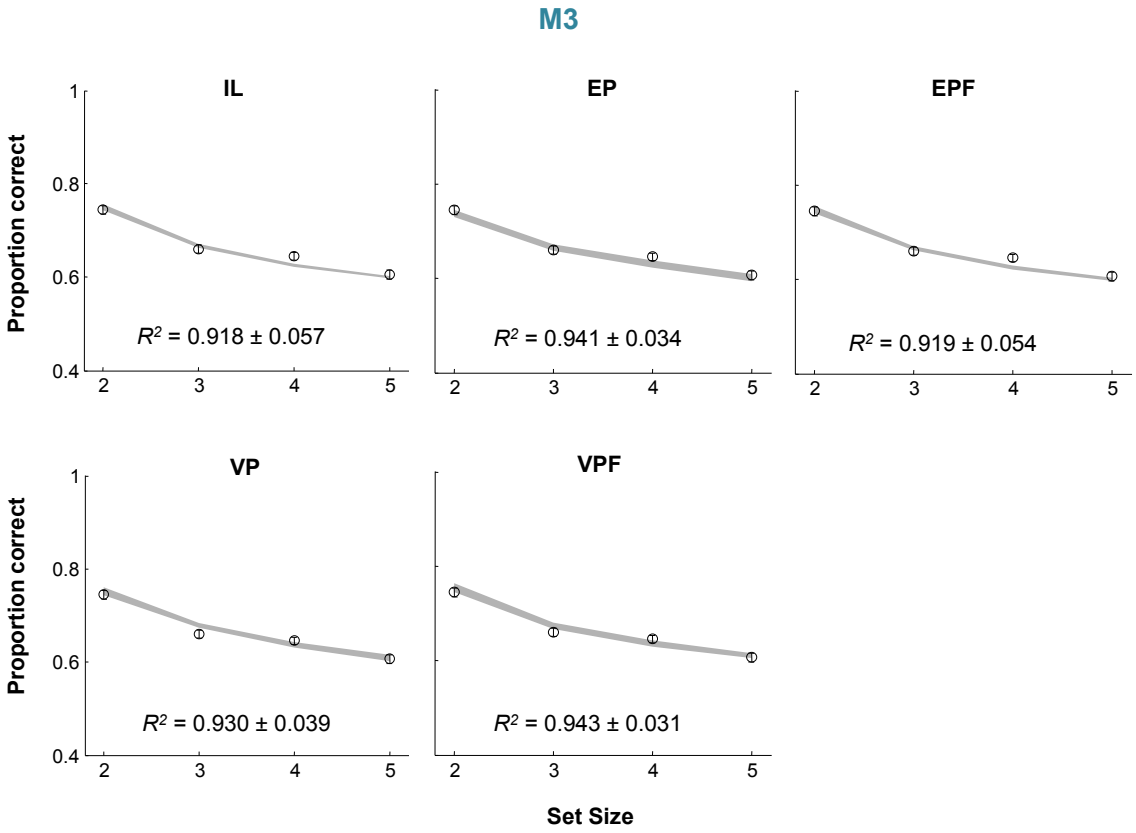


Figure 3.8. Proportion correct as a function of set size for M3. Circles and error bars: behavior; shaded areas: model fits. IL: Item-Limit, EP: Equal-Precision, EPF: Equal-Precision + Fixed Capacity, VP: Variable-Precision, VPF: Variable-Precision + Fixed Capacity.

The added manipulation of change magnitude, however, clearly separates these model predictions (Figures 3.9-3.12). The data from both species are best described by the two variable precision models, one in which there is a fixed capacity limit and another without a fixed capacity limit (VPF: R^2 values, M1: 0.902 ± 0.022; M2: 0.887 ± 0.024; M3: 0.896 ± 0.020; Humans: 0.803 ± 0.023; VP: M1: 0.90 ± 0.023; M2: 0.885 ± 0.024; M3: 0.891 ± 0.020; Humans: 0.799 ± 0.023), followed by the equal-precision + fixed-capacity model (R^2 values, M1: 0.755 ± 0.046; M2: 0.854 ± 0.030; M3: 0.817 ± 0.035; Humans: 0.714 ± 0.043), the equal-precision model (R^2 values, M1: 0.718 ± 0.055; M2: 0.835 ± 0.031; M3: 0.817 ± 0.035; Humans: 0.619 ± 0.037), and the item-limit model (R^2 values, M1: 0.402 ± 0.041; M2: 0.222 ± 0.038; M3: 0.263 ± 0.049; Humans: 0.228 ± 0.048).

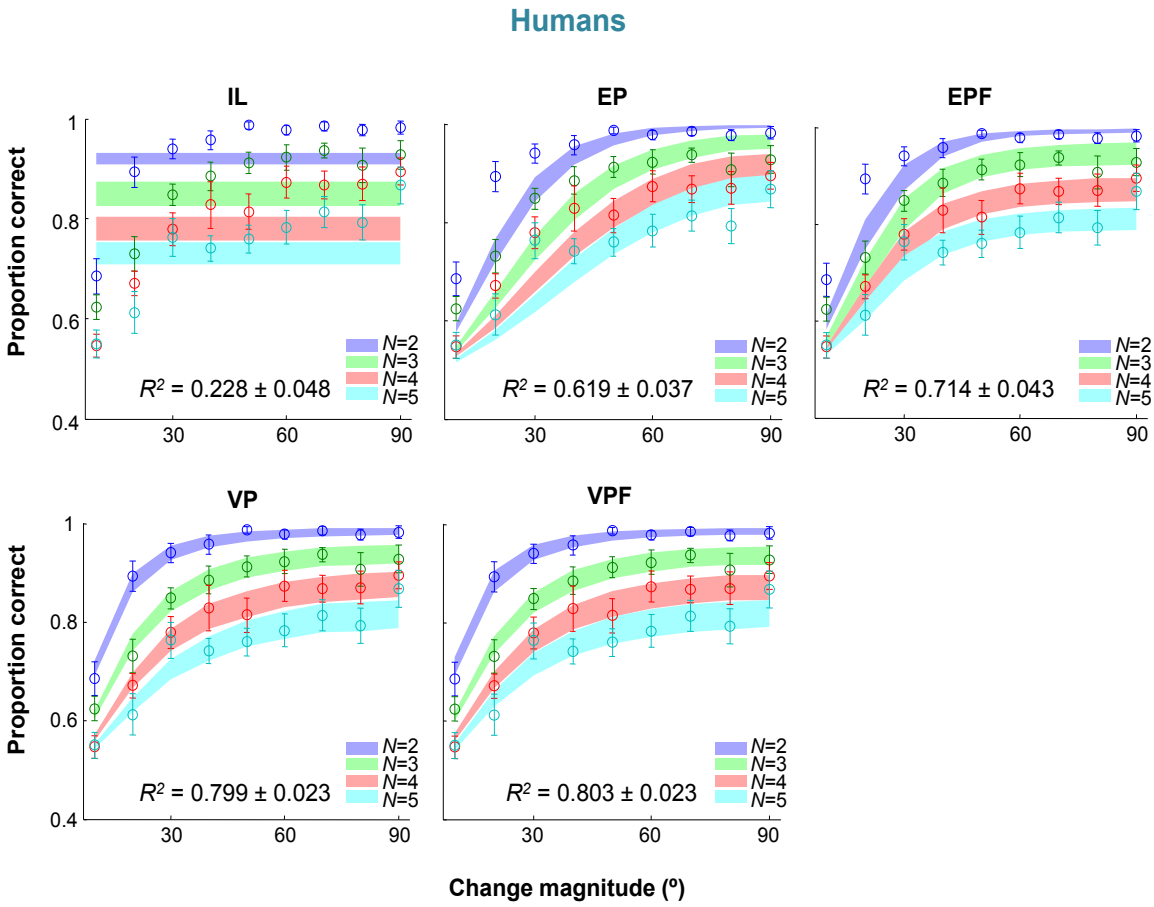


Figure 3.9. Proportion correct as a function of set size (*N*) and change magnitude (º) for humans. Circles and error bars: behavior; shaded areas: model fits (mean ± s.e.m. across subjects). IL: Item-Limit, EP: Equal-Precision, EPF: Equal-Precision + Fixed Capacity, VP: Variable-Precision, VPF: Variable-Precision + Fixed Capacity.

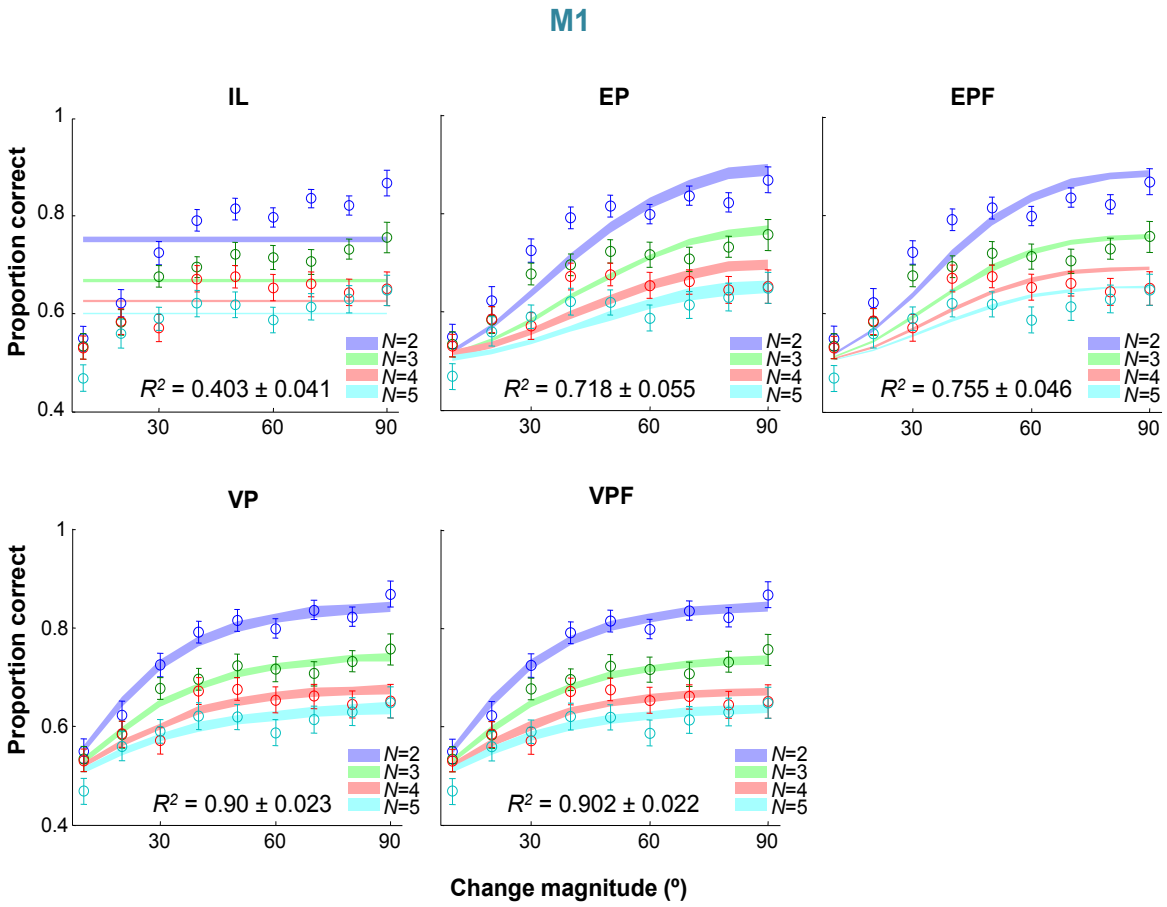


Figure 3.10. Proportion correct as a function of set size (*N*) and change magnitude (º) for M1. Circles and error bars: behavior; shaded areas: model fits (mean ± s.e.m. across bootstrapped datasets). IL: Item-Limit, EP: Equal-Precision, EPF: Equal-Precision + Fixed Capacity, VP: Variable-Precision, VPF: Variable-Precision + Fixed Capacity.

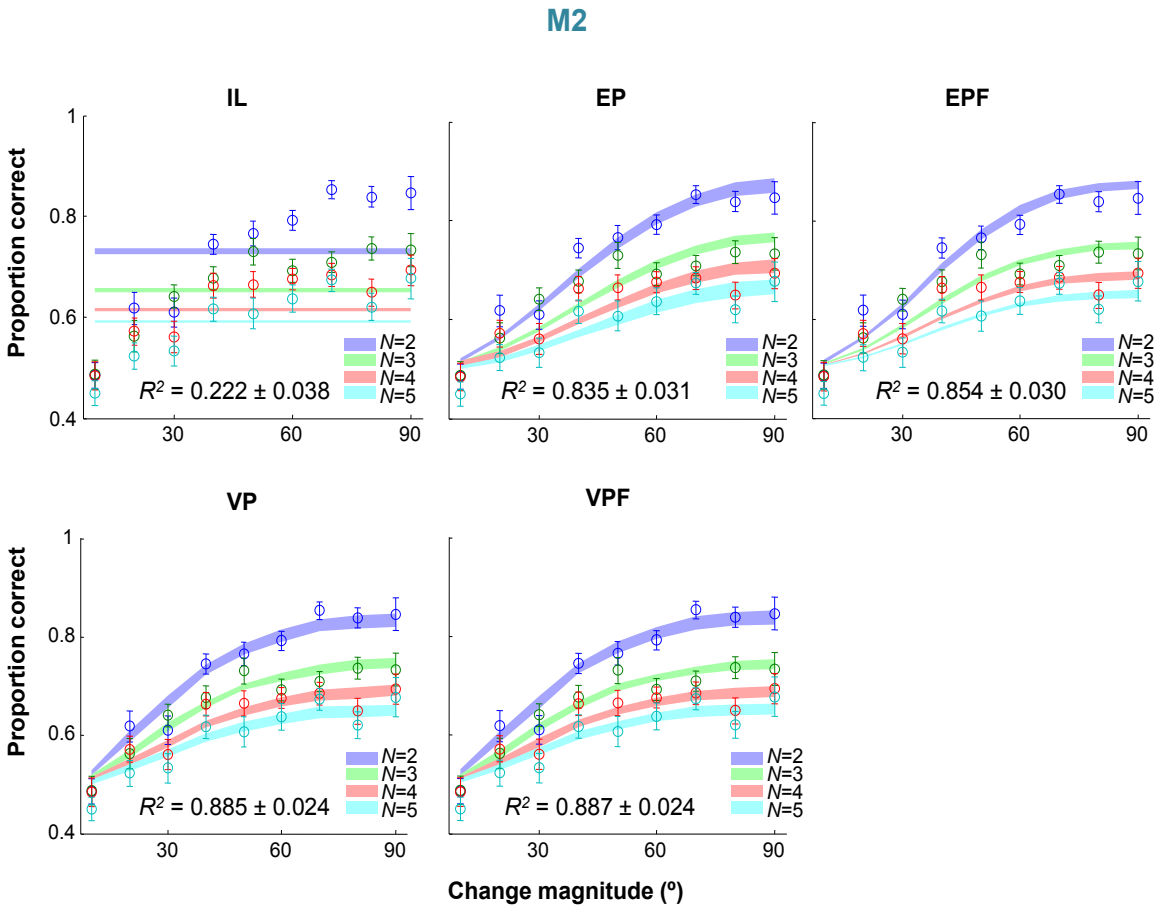


Figure 3.11. Proportion correct as a function of set size (*N*) and change magnitude (º) for M2. Circles and error bars: behavior; shaded areas: model fits (mean ± s.e.m. across bootstrapped datasets). IL: Item-Limit, EP: Equal-Precision, EPF: Equal-Precision + Fixed Capacity, VP: Variable-Precision, VPF: Variable-Precision + Fixed Capacity.

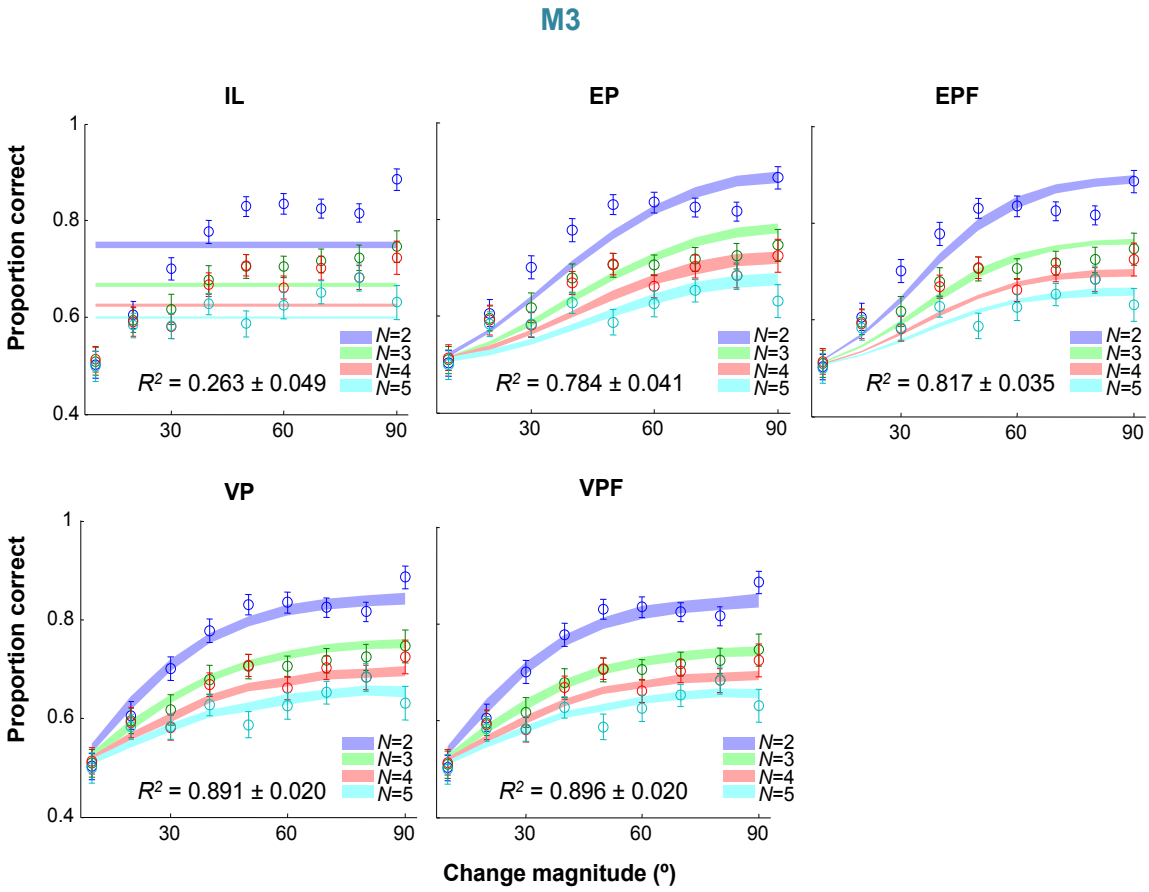


Figure 3.12. Proportion correct as a function of set size (*N*) and change magnitude (°) for M3. Circles and error bars: behavior; shaded areas: model fits (mean ± s.e.m. across bootstrapped datasets). IL: Item-Limit, EP: Equal-Precision, EPF: Equal-Precision + Fixed Capacity, VP: Variable-Precision, VPF: Variable-Precision + Fixed Capacity.

Possibly a more principled way to compare models is the Bayesian model comparison, a likelihood-based method that automatically corrects for the number of free parameters. It is important to note that the variable precision model with the fixed capacity limit is not a substantial improvement over the variable precision model without an upper bound, when taking into account that the former has an extra parameter (in humans, the latter actually performs marginally better). We thus prefer the simpler version of the variable precision model. We found that the log marginal likelihood of the variable-precision model exceeded that of the equal-precision + fixed capacity, equal-precision, and item-limit models for both species (Figures 3.13 - 3.16); this result remains unchanged when other model comparison metrics are used (see Table 2.1). These findings demonstrate that both monkey and human VWM are not limited by a fixed item capacity, but instead gradually deteriorate as more items have to be remembered. Despite the quantitative differences in memory performance between species, the success of the variable-precision model for both species demonstrates qualitative similarity and suggests evolutionary continuity of basic VWM mechanisms.

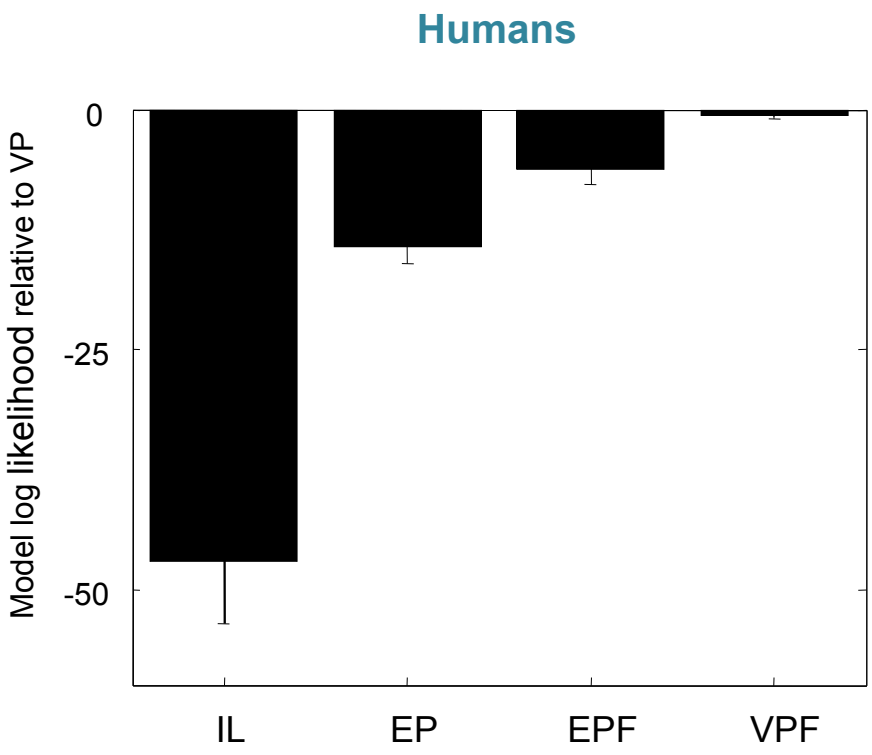


Figure 3.13. Marginal log likelihoods of the item limit, equal-precision, equal-precision with fixed capacity, and variable-precision with fixed capacity models minus those of the variable-precision model (mean ± s.e.m.), for humans. A value of $-x$ means that the data are e^x times more probable under the variable-precision model.

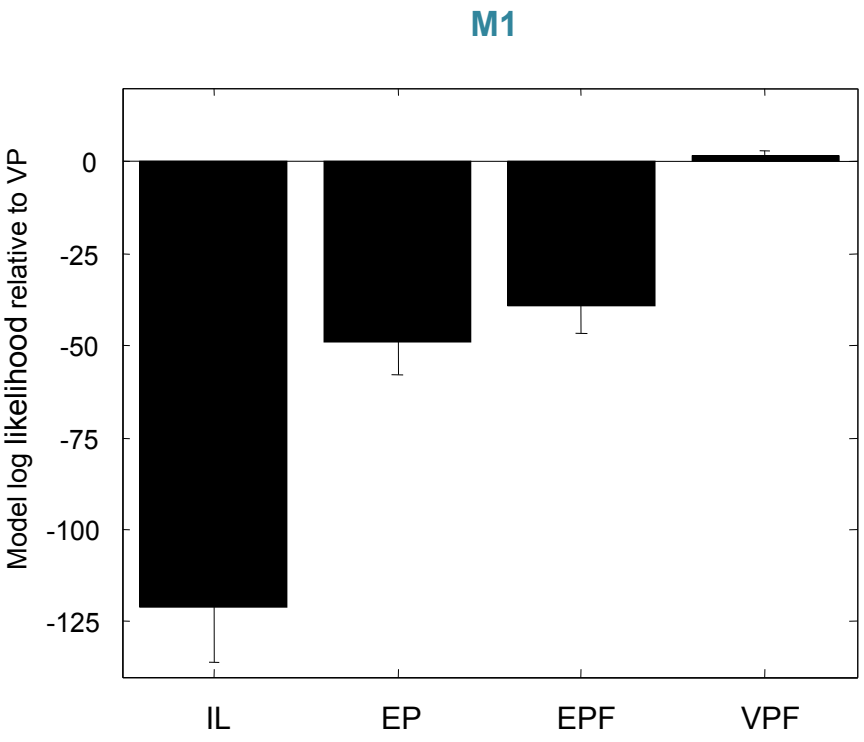


Figure 3.14. Marginal log likelihoods of the item limit, equal-precision, equal-precision with fixed capacity, and variable-precision with fixed capacity models minus those of the variable-precision model (mean ± s.e.m.), for M1. A value of $-x$ means that the data are e^x times more probable under the variable-precision model.

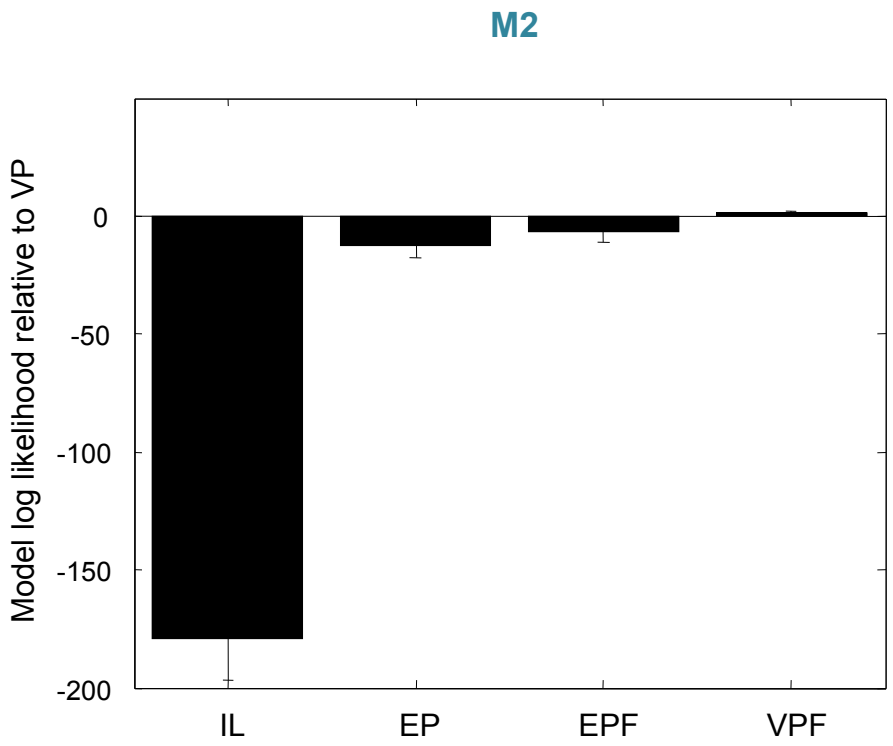


Figure 3.15. Marginal log likelihoods of the item limit, equal-precision, equal-precision with fixed capacity, and variable-precision with fixed capacity models minus those of the variable-precision model (mean ± s.e.m.), for M2. A value of $-x$ means that the data are e^x times more probable under the variable-precision model.

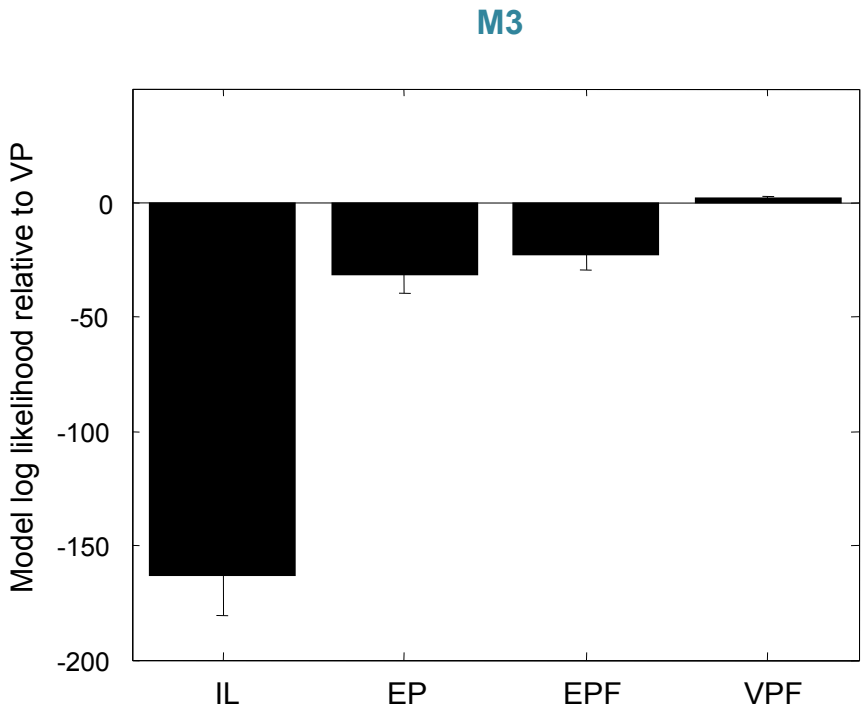


Figure 3.16. Marginal log likelihoods of the item limit, equal-precision, equal-precision with fixed capacity, and variable-precision with fixed capacity models minus those of the variable-precision model (mean ± s.e.m.), for M3. A value of $-x$ means that the data are e^x times more probable under the variable-precision model.

Model		AIC*(model)-AIC*(VP)	AICc*(model)-AICc*(VP)	BIC*(model)-BIC*(VP)		LML(model)-LML(VP)	
		Mean	Mean	Mean	s.e.m.	Mean	s.e.m
IL	M1	-126	-125	-122	15	-121	15
	M2	-184	-183	-180	18	-180	18
	M3	-168	-167	-164	18	-163	18
	Humans	-48.2	-47.2	-45.7	6.8	-47.1	6.6
EP	M1	-48.5	-47.5	-44.8	9.2	-48.9	9.1
	M2	-13.8	-12.8	-10.1	4.6	-12.7	4.8
	M3	-31.3	-30.3	-27.6	7.8	-31.3	8.1
	Humans	-13.9	-12.9	-11.4	1.5	-14.4	1.7
EPF	M1	-40.2	-40.2	-40.2	7.9	-39.0	7.8
	M2	-9.3	-9.3	-9.3	4.4	-6.7	4.6
	M3	-24.0	-24.0	-24.0	6.7	-22.6	6.9
	Humans	-7.6	-7.6	-7.6	1.5	-6.2	1.6
VPF	M1	-0.50	-1.3	-4.18	0.83	1.5	1.5
	M2	-0.31	-2.2	-4.0	0.91	1.2	0.81
	M3	0.44	-0.56	-3.2	1.5	2.0	1.1
	Humans	-0.48	-1.46	-3.00	0.32	-0.57	0.31

Table 3.1. Model comparisons of each model (IL: item-limit, EP: equal-precision, EPF: equal-precision + fixed capacity, and VPF: variable-precision + fixed capacity) showing mean differences minus that of the variable-precision model. The standard error of the mean is the same across AIC, AICc, and BIC because these measures have the same leading term, LL_{max}, and only differ in their penalty terms.

Given that the fundamental nature of VWM limitations is consistent between these two species, the quantitative performance differences may simply be due to differences in their parameter values within the same model (Tables 3.2 and 3.3). Mean precision, $\overline{J_1}$, was much lower in monkeys than in humans, which might be related to attentional differences between the two species. The exponent, α, in the relationship between mean precision and set size was similar across monkeys and somewhat higher in humans.

Humans

Model	Parameter	Tested range				
		Min	Step	Max	Mean	s.e.m.
IL	K	1	1	5	1.50	0.17
	ε	0	0.003	3	0.079	0.011
EP	J_1	0	0.13	25	17.7	2.3
	α	0	0.015	3	2.07	0.12
EPF	K	1	1	5	2.20	0.36
	J_1	0	0.13	25	15.3	2.5
	α	0	0.015	3	1.43	0.25
VP	$\overline{J_1}$	0	1.01	100	65.8	8.7
	τ	0.1	1.11	100	29.3	8.5
	α	0	0.03	3	1.82	0.13
VPF	$\overline{J_1}$	0	2.02	200	83.0	17.9
	τ	0.1	2.13	200	25.3	5.4
	α	0	0.061	3	1.97	0.18
	K	1	1	5	4.10	0.28

Table 3.2. Parameter ranges and parameter estimates for humans. Means and standard errors were computed across subjects.

Monkeys

Model	Parameter	Tested range			M1		M2		M3	
		Min	Step	Max	Mean	s.e.m.	Mean	s.e.m.	Mean	s.e.m.
IL	K	1	1	5	1	0	1	0	1	0
	ε	0	0.003	3	0.248	0.0055	0.269	0.0062	0.249	0.0065
EP	J_1	0	0.13	25	3.51	0.78	2.41	0.68	2.71	0.69
	α	0	0.015	3	2.35	0.21	1.98	0.27	1.98	0.26
EPF	K	1	1	5	1	0	1.16	0.79	1.12	0.62
	J_1	0	0.13	25	1.23	0.085	1.12	0.16	1.35	0.31
	α	0	0.015	3	1.35	0.60	1.57	0.74	1.89	0.83
VP	$\overline{J_1}$	0	0.30	30	11.0	1.8	3.82	0.87	7.0	1.8
	τ	0.1	0.40	30	24.9	4.4	6.2	2.6	15.7	5.9
	α	0	0.03	3	1.47	0.14	1.32	0.14	1.31	0.13
VPF	$\overline{J_1}$	0	0.30	30	10.2	2.7	3.7	1.4	7.7	2.7
	τ	0.1	0.40	30	23.8	5.1	5.6	2.8	13.9	5.2
	α	0	0.03	3	1.55	0.49	1.47	0.48	1.5	0.41
	K	1	1	5	3.6	1.8	3.4	1.5	3.2	1.1

Table 3.3. Parameter ranges and parameter estimates for monkeys. Means and standard errors were computed across bootstrapped datasets.

Our findings provide cross-species evidence that visual information is encoded in working memory in a noisy manner, with precision per item varying across items and trials, and on average decreasing with increasing set size. This is consistent with mounting evidence in humans (Keshvari et al., 2012, 2013; van den Berg et al., 2014; van den Berg, Shin, et al., 2012). Variability in precision could result from a variety of factors including, eye movements, interference from other stimuli, and fluctuations in attention. At the neural level, precision may correspond to the gain of a neural population pattern encoding the stimulus (Ma et al., 2014; van den Berg, Shin, et al., 2012). Neurophysiological evidence supports this notion, showing that firing rate decreases as set size increases (Churchland, Kiani, & Shadlen, 2008). A plausible mechanistic implementation of the variable-precision model was recently proposed (Bays, 2014). Thus, at present, behavioral, physiological, and computational data seem to unambiguously point toward resource models as the best account of VWM limitations, and our results establish rhesus monkeys as a suitable model system for further elucidating the neural substrates of these limitations.

CHAPTER 4: CHANGE DETECTION TESTING IN RHESUS MONKEYS AND HUMANS: DECISION-MAKING

Introduction

One of the key functions of the brain is to process and interpret sensory information to make perceptual decisions. Sensory information available to the subject is necessarily limited. For example, the sensory measurements might be of low quality, due to both internal (noisy encoding) and external (e.g., poor contrast, distant or fast-moving objects, etc.) factors. Even when sensory quality is high, the same sensory stimulus is subject to multiple interpretations. Given this partially informative sensory information, our inference is probabilistic; that is, it comes with some degree of uncertainty. The brain must evaluate the uncertain sensory information effectively, make judgments relative to goals, and respond to the environment accordingly.

As described in previous chapters, the precision with which sensory information is represented internally can vary across items and trials, and in many perceptual tasks, the degree of precision not only has a role in the encoding stage, but also plays a role in the observer's decision-making stage. The change detection task taps into both of these stages of VWM processing. The observer encodes and maintains information about the sample stimuli, compares its memory of the sample stimuli with the test stimuli at the corresponding locations, makes a judgment about which test stimulus has changed, and then makes a response. In the change-detection task, the precision of the task-relevant stimulus can be manipulated by changing the reliability of the stimulus; for example, the height to width ratio of an ellipse in an orientation change detection task, such that a low reliability ellipse would be

shorter and wider, making the orientation change more difficult to discriminate. The question, then, is how the observer factors in unreliable stimuli along with reliable stimuli in making judgments and decisions about which stimulus changed in orientation.

A Bayesian-optimal observer would learn to use the noisy internal representations of the stimuli and essentially compute a probability distribution indicating where the change occurred, giving more weight to reliable evidence (e.g., longer and thinner ellipses) than to unreliable evidence (e.g., shorter, fatter ellipses). Based on the computed relative weight of evidence, the observer chooses the location with the higher probability of change. Since the uncertainty can change on an item-by-item and trial-by-trial basis, this requires trial-to-trial computations of probability distributions over stimulus features. These types of computations are referred to as probabilistic computations (see Ma, 2012 for a review).

In a related visual change-detection paradigm, human subjects were asked to identify whether or not an orientation change occurred in the test display containing four ellipses (Keshvari, Van den Berg, and Ma, 2012). Reliability of the stimuli was manipulated (as previously described) and the results were evaluated by fitting an optimal-observer model against suboptimal models to assess how the observers took into account the reliability of the stimuli and the variability in precision. The decision models that were tested differed according to the assumption that the observer makes about encoding precision.

Performance was best described by an optimal model, in which the observer had complete knowledge about precision for every item on every trial. In the decision stage, the observer uses this information about encoding precision and about the reliability stimulus differences on an item-by-item and trial-by-trial basis to make an optimal decision about change. Two suboptimal models were also tested in which the observer makes an incorrect assumption about precision by either 1) assuming that precision is completely determined by the reliability of the stimulus and ignored other sources of variability or 2) assuming a single value of precision throughout the experiment, and ignored all sources of variability including manipulations of reliability of the stimulus.

Although monkeys and humans were shown to have similar VWM encoding mechanisms (Chapter 3), it is nevertheless an open question whether monkeys would employ similar or different decision processes in this change detection task. To this end, two rhesus monkeys and ten humans were tested to determine similarities or differences in decision processes in nearly identical change detection tasks.

Subjects briefly viewed a sample array of three randomly oriented ellipses (fixed set size) and, following a delay, a test array containing two randomly chosen ellipses from the sample array, of which one had a different orientation than in the sample array (Figure 4.1). The reliability of the stimulus was manipulated by changing the height-to-width ratio of the ellipse (see Methods). Thus, on each trial, each item could either be a high-reliability ellipse (longer, thinner) or a low-reliability (shorter, wider) ellipse. Furthermore, in a manner

similar to Aim 1, the magnitude of the orientation change could take one of nine values and subjects identified which item had changed orientation by touching it, and received trial-to-trial feedback.

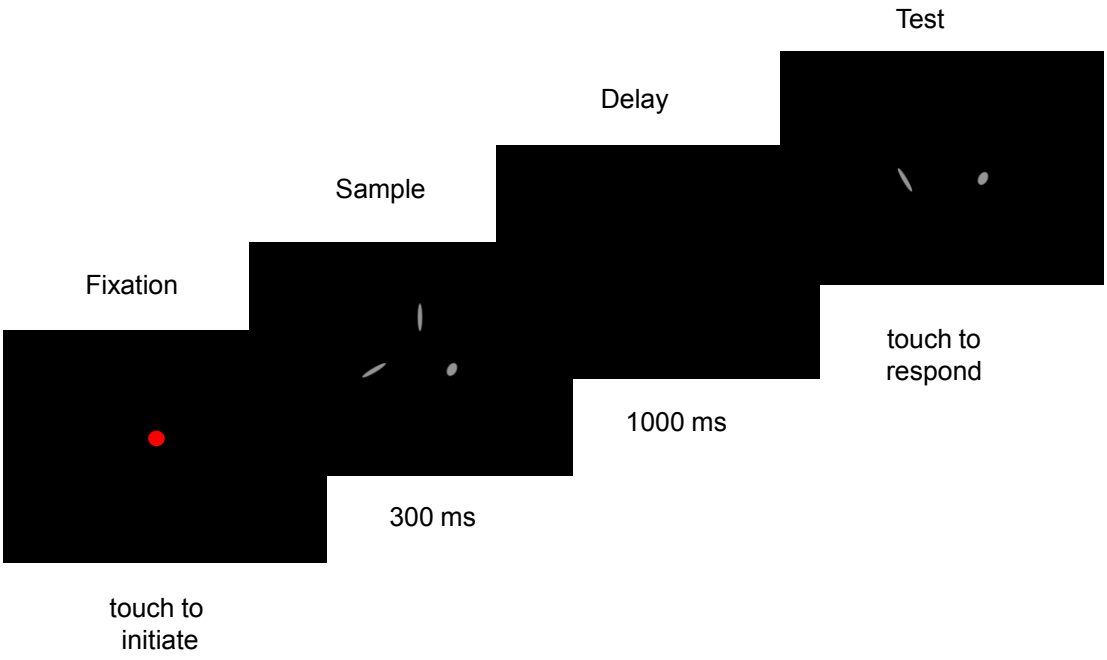


Figure 4.1. Task Procedure. Schematic representation showing a sample trial. Set size was always 3. Stimulus reliability was controlled by ellipse elongation.

The decision models tested by Keshvari et al. (2012) were adapted to fit the somewhat different change-detection task used here where subjects were presented with only two test stimuli each trial and had to report which changed. The three decision models considered here differed only in the decision rule. For the encoding stage, precision was modeled based on the variable-precision model (Chapter 3) since evidence from the first experiment suggested variable precision encoding by both species (Figure 4.2). The decision rules are based on the assumption that the observer makes about encoding precision.

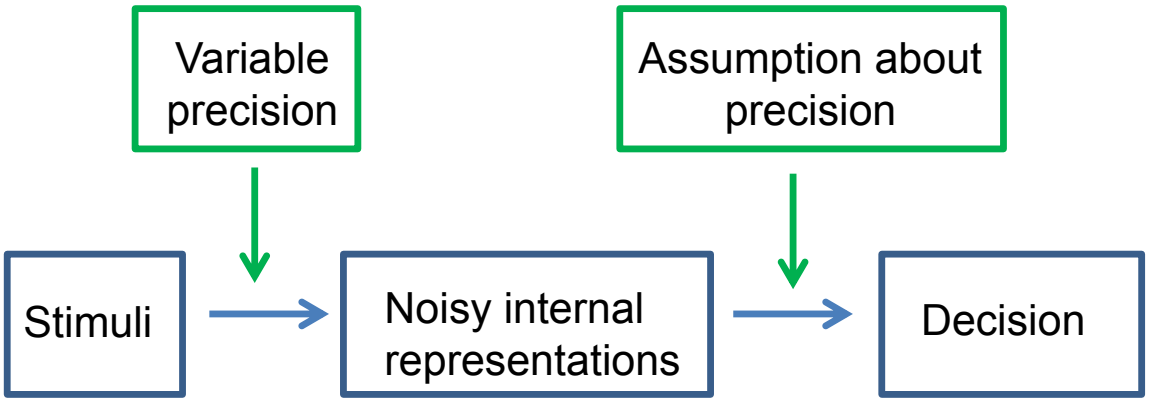


Figure 4.2. Flow diagram of the encoding and decision processes. Models here are identical in the encoding stage, but differ in the assumption that the observer makes about precision.

The theory and derivations of these models are described in the Methods section. In the names of the models below, the “variable precision” before the hyphen indicates encoding stage and the latter indicates the decision stage, which is based on the observer’s assumption about precision.

1) Variable Precision-Variable Precision (Optimal): Encoding precision is variable and the observer has complete knowledge of an item’s precision on each trial. The observer, thus, computes precision on a trial-by-trial and item-by-item basis, taking into account the reliability of the stimulus and giving more weight to more reliable stimuli.

2) Variable Precision-Equal Precision (Suboptimal): Encoding precision is variable; however, the observer assumes that precision is completely determined by the reliability of the stimulus, and ignores any other variability in encoding precision across items and trials. The observer, thus, assumes only two values of precision: high (when the stimulus has high reliability) and low (when the stimulus has low reliability).

3) Variable Precision-Single Precision (Suboptimal): Encoding precision is variable; however, the observer assumes that precision stays constant throughout the experiment; thus, ignoring both variations due to reliability of the stimuli and other variability. The observer, then, applies a single value of precision across all items and trials.

Methods

Monkeys

Subjects

Two adult male rhesus monkeys, M1 and M2 from Chapter 3 participated in this experiment (M3 could not participate because of health issues) five days each week. All animal procedures were performed in accordance with the National Institutes of Health guidelines, approved by the Institutional Review Board at University of Texas Health Science Center at Houston, and supervised by the Institutional Animal Care and Use Committee.

Apparatus

Monkeys were tested with the same apparatus as the one described in Chapter 3.

Stimuli

Stimuli consisted of white ellipses displayed on a black background. Two types of ellipses were used: “high-reliability” (long and narrow) and “low-reliability” (short and wide). In this experiment, ellipses were chosen instead of oriented bars (ones used in Chapter 3) because ellipses contain less corner information and orientation changes are more difficult to discriminate, particularly as the ellipses get shorter and wider. The size of the high-reliability stimuli was 1.8 x 0.4 cm and that of the low-reliability stimuli was 0.93 x 0.77 cm. The area of both types of stimuli remained the same; only the height-to-width ratio was changed. Based on the average distance of the monkey from the screen

(approximately 35 cm), the high-reliability and low-reliability stimuli subtended a visual angle of 2.9° x 0.65 ° and 1.5° x 1.3° respectively. Stimuli were presented in six possible locations on the screen, arranged on an imaginary circle (same as Chapter 3).

Trial procedure

Each trial began with a red fixation point in the center of the screen as shown in Figure 4.1. The monkeys had to make a one-touch response to the fixation point, which initiated the presentation of a sample display. In this experiment, the set size remained fixed, so the sample display always contained three items, and had a presentation time of 300 ms. After a delay of 1000 ms, the test display was presented, which always consisted of two items, placed at the same locations as two randomly chosen items from the sample display. One test item had the same orientation as the corresponding item in the sample display, and the other test item had a different orientation. It is important to note that the test items always had the same stimulus type (reliability) as the sample items. For example, a high-reliability sample item would always be a high-reliability if it was chosen as a test item; the only change occurred in the orientation of the changed item. The monkeys' task was to identify which item had changed, and to touch that item. The test display remained on the screen until response. Correct responses were rewarded. An intertrial interval of 3000 ms followed the choice response, during which a green light illuminated the chamber and the screen was dark. There were four trial conditions: 1) Both test items were of high reliability; therefore the changed item was of high reliability 2)

The changed item was of high reliability and the unchanged item was of low reliability 3) The changed item was low-reliability and the unchanged item was high-reliability 4) Both test items were low-reliability, so the changed item was low-reliability.

Training

Both monkeys that participated in this study (M1 and M2) had been previously trained in a change detection task using oriented bars. For preliminary training of these two monkeys, we intermixed trials of oriented bars (old stimuli) with trials of oriented ellipses for initial task acquisition, which required 6 sessions per monkey. Once the monkeys' performance on ellipse trials was similar to baseline performance with oriented bars, we began training them with only ellipses trials. Both monkeys were first trained with only high-reliability ellipses at a set size of 2 and a sample viewing time of 300 ms. Once overall accuracy reached approximately 70%, the monkeys were tested with a set size of 3. Finally, we gradually intermixed the low-reliability trials with set size 3 and once the monkeys' performance on these trials reached 60%, they were ready for testing. For M1 and M2, this training required 28 and 32 sessions respectively.

Testing

Set size was fixed at 3. On every trial, each item had an equal probability of being a high- or low-reliability ellipse. The locations of the ellipses were chosen randomly from 6 possibilities. The orientation of each sample item, θ, was drawn independently from a uniform distribution over 18 possible

orientations (−90°, −80°,…, −10°, 10°, 20°,…, 80°). The orientation of the changed item in the test display was drawn from the same uniform distribution. Testing consisted of 60 sessions, with 192-trial blocks per session, for a total of 11,520 trials per monkey.

Humans

Subjects

Ten human subjects (8 females) aged 21-33 years (mean age = 27.1 years) participated. Each subject visited the lab for two 1.5-hour sessions and was compensated $10 per session. Study procedures were approved by the University of Texas Health Science Center at Houston Institutional Review Board.

Apparatus and Stimuli

The apparatus for this experiment remained the same as the one used to test subjects in Experiment 1 (Chapter 3). Subjects were seated in a chair in a small room equipped with a computer. At the beginning of the experiment, the distance between the chair and the screen was adjusted so that the stimuli and display would subtend approximately the same visual angles as for the monkeys. Subjects were asked to maintain approximately the same distance. The monitor and touchscreen were identical to those used for monkeys.

Trial Procedure

The trial procedure was identical to that for the monkeys, except for the feedback. Feedback consisted of a green light that was illuminated for 1 s and accompanied by a tone for correct responses, or a red light illuminated for 1 s for incorrect responses (same feedback as the one in Chapter 3 experiment).

Training and Testing

Each subject completed two testing sessions, each consisting of three 192-trial blocks, for a total of 1152 trials per subject. Subjects were given a 10-minute break time in between blocks. Each subject completed 8 practice trials at the beginning of the first session (same from Chapter 3 experiment).

Theory and Modeling

Three models of decision-making were considered in this change-detection task. On a given trial, the observer's decision process consists of an encoding stage and a decision stage. Here, the encoding stage in all three models remained identical and was modeled according to the variable precision model from Aim 1 (see below).

Encoding stage: Variable Precision

For this task, we model encoding precision as a random variable. The theory and derivations of this encoding stage remain largely the same as one described in Chapter 2. First, as in Experiment 1, the orientation space was mapped to the interval $[0,2\pi)$ by multiplying all orientations and orientation

change magnitudes by 2 before analysis for simplification. This method was expressed in all equations shown below; however, for the figures, the change magnitudes were mapped back to actual orientation space. Second, it was assumed that both orientations in the test display, denoted by φ_1 and φ_2, are known noiselessly to the observer, because they remain on the screen until the observer responds. Third, the relationship between encoding precision (expressed as Fisher information denoted J) and the concentration parameter (denoted κ) remains the same as in Chapter 2:

$$J = \kappa \frac{I_1(\kappa)}{I_0(\kappa)}, \qquad (6)$$

where I_1 is the modified Bessel functions of the first kind of order 1.

As previously described for the variable precision model, encoding precision is a random variable that follows, independently for each item and each trial, a gamma distribution with mean $\bar{J}$ and scale parameter τ, denoted $p(J \mid \bar{J};\tau)$ with mean $\bar{J}$ and variance $\bar{J}\tau$.

A key difference in this experiment is that mean $\bar{J}$ differs between the stimulus type: encoding precision is drawn from two distributions depending upon stimulus type: For a high-reliability item, mean precision is $\bar{J}_{high}$. For a low-reliability item, mean precision is $\bar{J}_{low}$. The noisy measurements x_1 and x_2 are drawn from a doubly stochastic process, where first, precisions J_1 and J_2 are each drawn from a gamma distribution with mean $\bar{J}_{high}$ or $\bar{J}_{low}$ and scale

parameter τ; then, the noisy measurements are drawn from the Von Mises distribution (Eqn 2) with concentration parameters κ_1 and κ_2, which are determined by J_1 and J_2 respectively (Eqn 1).

$$p(x_i \mid \theta_i) = \frac{1}{2\pi I_0(\kappa_i)} e^{\kappa_i \cos(x_i - \theta_i)}. \qquad (2)$$

where I_0 is the modified Bessel functions of the first kind of order 0 and the noisy memory of the i^{th} item in the sample display, denoted x_i.

Generative model and Inference

The generative model and inference in this experiment remains the same as described in Chapter 2. The resulting decision rule is identical to the one used to model the decision process in Experiment 1. For derivations of the log posterior ratio, refer to Chapter 2.

The ideal observer responds that the change occurred at location 1 when the log posterior ratio is positive:

$$\log \frac{I_0(\kappa_1)}{I_0(\kappa_2)} + \kappa_2 \cos(x_2 - \varphi_2) - \kappa_1 \cos(x_1 - \varphi_1) > 0$$

This decision rule is valid for all decision models considered here. They simply differ in how the concentration parameters, κ_1 and κ_2 weigh the reliability of the stimulus. These differences are described below.

Model Differences in Decision Rules

In the Variable Precision-Variable Precision (optimal) model, the observer has complete knowledge of precision on every item and trial so uses the actual values of κ_1 and κ_2. The decision rule is then to localize the change at location 1 if

$$\log\frac{I_0(\kappa_1)}{I_0(\kappa_2)} + \kappa_2\cos(x_2-\varphi_2) - \kappa_1\cos(x_1-\varphi_1) > 0$$

In the Variable Precision-Equal Precision (suboptimal) model, the observer assumes equal precision for every item, on every trial. The assumed value is $\bar{J}_{\text{high}}$ for a high-reliability item, and $\bar{J}_{\text{low}}$ for a low-reliability item. These correspond to concentration parameters κ_{high} and κ_{low}. The decision rule, then, is the same as above but with κ_1 and κ_2 each taking on one of only two possible values, κ_{high} and κ_{low}, depending on the reliability of that item.

In the Variable Precision-Single Precision (suboptimal) model, the observer completely disregards the reliability of the stimulus and pretends $\kappa_1=\kappa_2$, so they do not use a weighted precision value. Then, the decision rule simplifies to:

$$\cos(x_2-\varphi_2) - \cos(x_1-\varphi_1) > 0$$

Lapse rate: For each of these models, we fitted a lapse rate parameter which accounts for errors due to lapses in attention, blinking or eye movements during stimulus presentation, or errors in making a response.

Each of these models have 4 parameters: $\bar{J}_{\text{high}}$, $\bar{J}_{\text{low}}$, τ, and lapse.

Model predictions

The probability of correct predicted by each model was computed for each stimulus condition, given the model parameters. Based on the assumptions in the generative model, in this task, the stimulus condition is determined by change magnitude Δ, and four condition types, *C*, described above: 1) Both high-reliability 2) Mix stimuli; high-reliability change 3) Mix stimuli; low-reliability change and 4) Both low-reliability. Thus, the probability that the decision rules for each of the models returns the correct location is computed when the memories x_1 and x_2 follow their respective distributions given *C* and Δ.

Each of these proportions correct was determined through Monte Carlo simulation, i.e. through a large number (10,000) of random draws of x_1 and x_2 and of J_1 and J_2. For each draw, the decision rule was evaluated, and then computed across all draws of the proportion of correct responses.

Finally, for each model, the parameter space was discretized finely (see Table 1) and a look-up table was calculated in which each entry gave the predicted probability of a correct response at one trial combination (*C*, Δ) for one parameter combination.

Model fitting

For a given model, the model parameters are denoted by a vector **t**. The log likelihood of **t** (the parameter log likelihood) is

$$LL(\mathbf{t}) = \log p(\text{data} \mid \text{model}, \mathbf{t}) = \log \prod_{i=1}^{n_{\text{trials}}} p(\text{correctness}_i \mid C_i, \Delta_i, \mathbf{t})$$

where the product is over trials (from 1 to n_{trials}) and correctness$_i$ is 1 if the subject was correct on the i^{th} trial and 0 if not. We can rewrite this as

$$\begin{aligned} LL(\mathbf{t}) &= \sum_{i=1}^{n_{\text{trials}}} \log p(\text{correctness}_i \mid C_i, \Delta_i, \mathbf{t}) \\ &= \sum_{C} \sum_{\Delta} \begin{bmatrix} n(C, \Delta, \text{correct}) \cdot \log p(\text{correct} \mid C, \Delta, \mathbf{t}) + \\ + n(C, \Delta, \text{incorrect}) \cdot \log(1 - p(\text{correct} \mid C, \Delta, \mathbf{t})) \end{bmatrix}, \end{aligned} \quad (4)$$

where trials are grouped by trial condition, C, change magnitude Δ, and by whether the observer was correct or incorrect, and n (C,Δ,correct) is the number of trials with a particular C, Δ, and correctness.

The same method of maximum likelihood estimation (as one described in Chapter 3) was used to compute parameter estimates. Thus, for each subject's data set Eq. (4) and the precomputed look-up table of model predictions mentioned above was used to find the log likelihood of each parameter combination. The parameter combination on this grid that maximized the log likelihood gives the estimates of the parameters. The model predictions corresponding to that parameter combination were then used to compute the model fits to the psychometric curves.

Model comparison

To compare models, the same four metrics from Chapter 3 were used: the Akaike Information Criterion (Akaike, 1974), the Akaike Information Corrected

Criterion (Burnham, 2002; Hurvich & Tsai, 1989), the Bayesian Information Criterion (Schwarz, 1978), and the log marginal likelihood (MacKay, 2003).

Bootstrapping

The original data set for each of the two monkeys consisted of 11,520 trials. The same bootstrapping method (described in Chapter 2) was used to create 100 bootstrapped data sets for each monkey. A random sample 11,520 trials (a combination of condition, change magnitude, and correctness) was selected with replacement from the original dataset to create each bootstrapped data set. The parameter estimates, psychometric curves, R^2 values, and model comparisons (AIC, AICc, BIC, and LML) were generated for each bootstrapped data set separately. The means for each of these were computed by averaging across all bootstrapped data sets from the same monkey, and the standard deviations served as estimates of the standard errors of the means.

Results and Discussion

Humans and monkeys both showed the highest proportion correct at condition 1, where both stimuli were high-reliability ellipses and the lowest proportion correct at condition 4, where both stimuli were low-reliability ellipses (Figure 4.3). Proportion correct was intermediate in the mix conditions (where one test item was high- and one was low-reliability) compared with conditions 1and 4. Interestingly, in the mix conditions, when the changed item was of high-

reliability, accuracy was higher than when the changed item was of low-reliability reflecting their training to look for and chose the stimulus that changed.

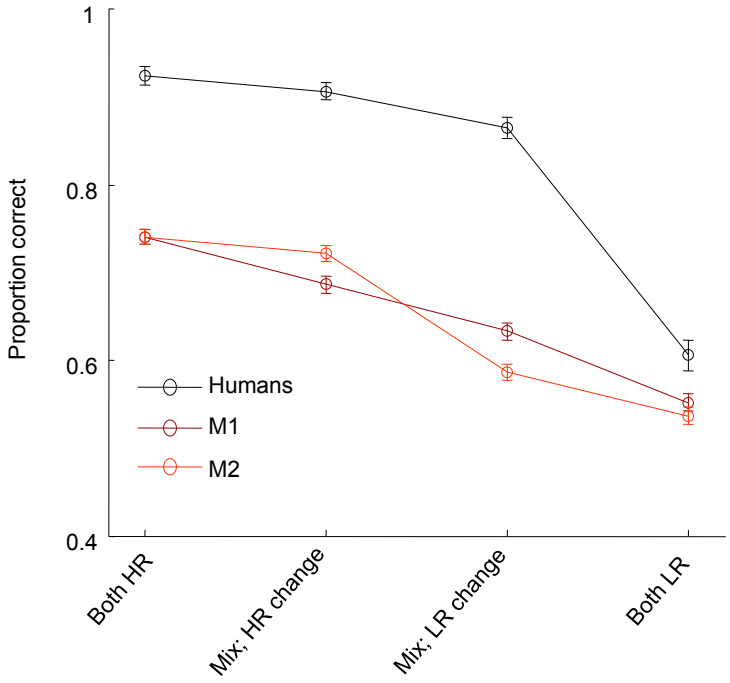


Figure 4.3. Proportion correct across four trial conditions for humans and two monkeys.

A more detailed representation of the data is provided by proportion correct as a function of change magnitude for each of the four conditions (Figure 4.4). We found large effects of both condition (high vs. low reliability) and change magnitude on performance for both species (humans: two-way repeated-measures ANOVA; condition: $F(3,27) = 231.35$, $p < 0.001$, change magnitude: $F(8,72) = 64.91$, $p < 0.001$; interaction: $F(24,216) = 15.73$, $p < 0.001$). Interestingly, the effect of change magnitude dissipates in both species in conditions 3 and 4, when the changed item is a low-reliability stimulus. This finding of change-magnitude dissipation in condition 3 suggests that both species took into account the uncertainty of the low-reliable stimulus and gave more weight to the high-reliable stimulus. And to the extent that the subjects could judge that the high-reliability item did not change, they were able to infer that the changed item then must be the low-reliability item. Thus, the change magnitude of the low-reliability item was less important since the decision could be based on a 'default' response when the subject was confident that the high-reliability item did not change. Overall, humans performed better than monkeys at all conditions; however, both species showed strikingly similar qualitative patterns of performance in all conditions.

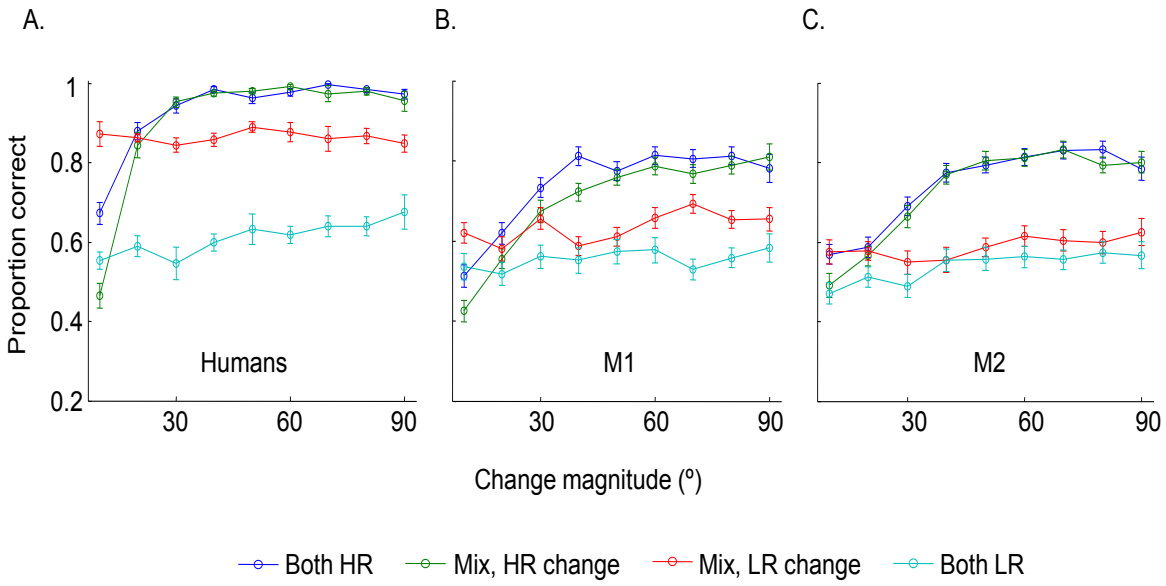


Figure 4.4. (A) Proportion correct as a function change magnitude across four trial conditions for humans (mean ± s.e.m. across ten subjects) (B) Same for M1 (mean ± s.e.m. across bootstrapped datasets) (C) Same for M2 (mean ± s.e.m. across bootstrapped datasets)

In spite of the large quantitative differences between the performances of both species, it is possible that the underlying mechanisms of perceptual decision-making are similar. To test this possibility, three decision models were compared for each individual monkey and human. The parameter estimates in each model were derived by using maximum-likelihood estimation and were fitted for each human subject as well as for each data set sampled using bootstrapping from an individual monkey's data.

Performance of both species was best described by the variable precision-equal precision suboptimal model (Humans: R^2 = 0.875 ± 0.018, M1: R^2 = 0.845 ± 0.025, M2: R^2 = 0.901 ± 0.021), followed by the variable precision-variable precision optimal model (Humans: R^2 = 0.814 ± 0.025, M1: R^2 = 0.782 ± 0.031, M2: R^2 = 0.641 ± 0.033), and the variable precision-single precision suboptimal model (Humans: R^2 = 0.431 ± 0.039, M1: R^2 = 0.507 ± 0.047, M2: R^2 = 0.391 ± 0.038; Figures 4.5 - 4.7). These conclusions were similar when likelihood-based model comparison metrics were used (Table 4.1).

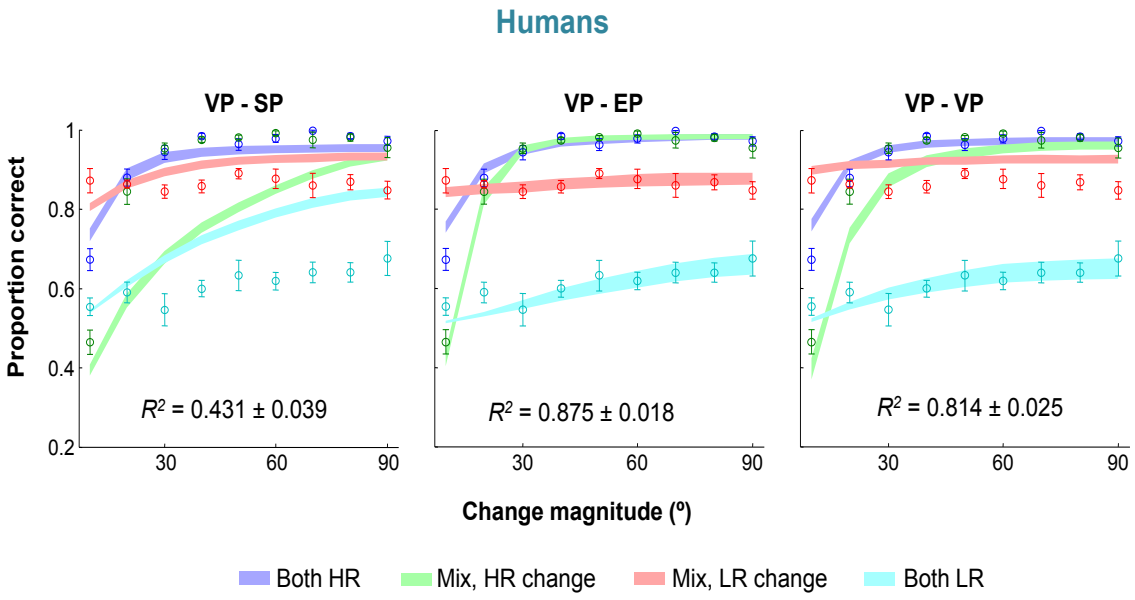


Figure 4.5. Proportion correct across trial conditions and change magnitudes (°) for humans. Circles and error bars: behavior; shaded areas: model fits (mean ± s.e.m. across subjects). VP-SP: Variable Precision-Single Precision, VP-EP: Variable Precision-Equal Precision, and VP-VP: Variable Precision-Variable Precision.

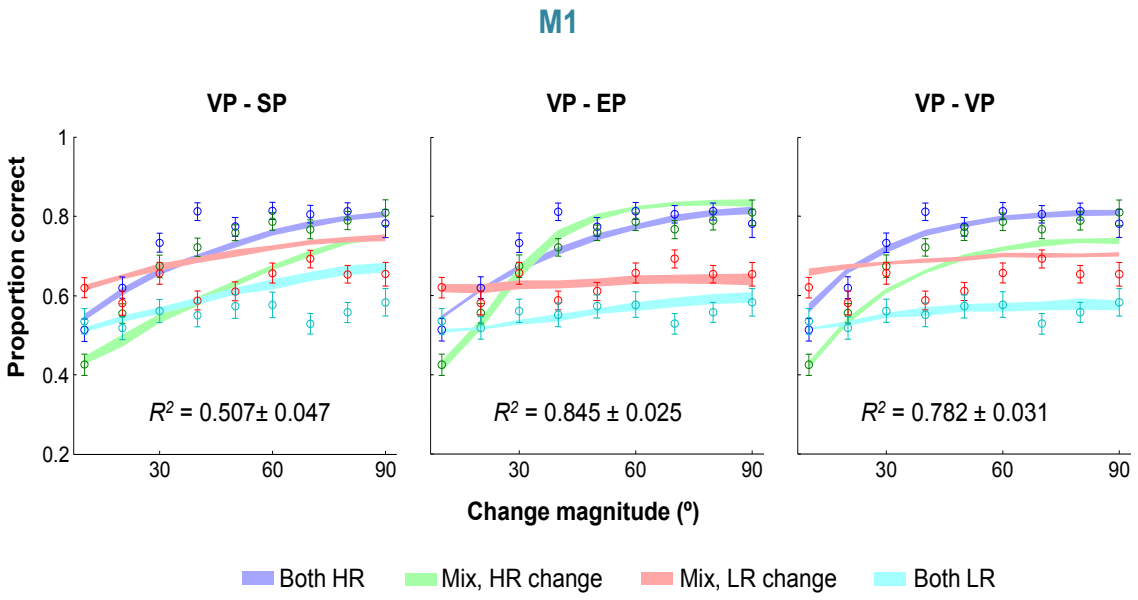


Figure 4.6. Proportion correct across trial conditions and change magnitudes (º) for M1. Circles and error bars: behavior; shaded areas: model fits (mean ± s.e.m. across bootstrapped datasets). VP-SP: Variable Precision-Single Precision, VP-EP: Variable Precision-Equal Precision, and VP-VP: Variable Precision-Variable Precision.

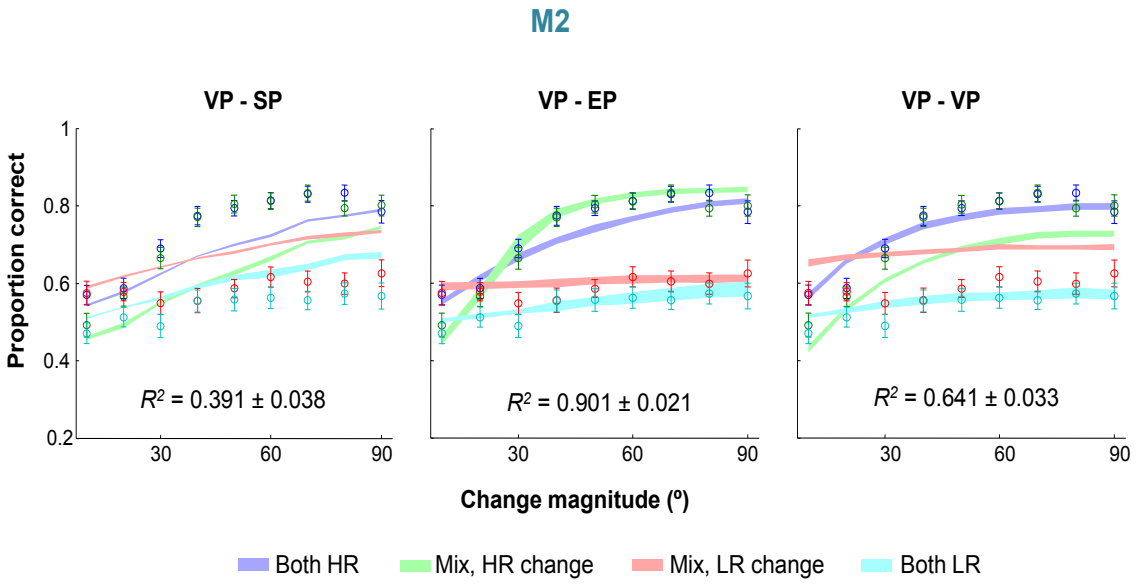


Figure 4.7. Proportion correct across trial conditions and change magnitudes (°) for M2. Circles and error bars: behavior; shaded areas: model fits (mean ± s.e.m. across bootstrapped datasets). VP-SP: Variable Precision-Single Precision, VP-EP: Variable Precision-Equal Precision, and VP-VP: Variable Precision-Variable Precision.

Model		AIC*/AICc*/BIC*(model)-AIC*/AICc*/BIC* (VP-EP)		LML(model)-LML(VP-EP)	
		Mean	s.e.m.	Mean	s.e.m
VP-SP	M1	-107	14	-107	14
	M2	-199	23	-204	24
	Humans	-57.5	5.9	-39	4.5
VP-VP	M1	-16	11	-19	11
	M2	-102	17	-107	17
	Humans	-12.1	2.4	-2.4	1.4

Table 4.1. Model comparisons of two models (VP-VP: variable precision-variable precision and VP-SP: variable precision-variable precision) showing mean differences minus that of the winning model, variable precision-equal precision model. The VP-EP model outperforms the other models according to all metrics. The means and standard error of the means are the same across AIC, AICc, and BIC because these measures have the same leading term, $LL_{\max}$, and an equal number of parameters.

These findings demonstrate that monkeys and humans are Bayesian-observers that take into account the uncertainty of sensory observations when making perceptual judgments, giving more weight to more certain evidence (see parameter estimates in Table 4.2). Even though encoding precision varies on an item-by-item and trial-by-trial basis, in this task, humans and monkeys make a wrong assumption about precision, and assume two values of precision that are completely determined by the reliability of the stimulus. They ignore other variability in encoding precision that may arise due to noise from both external and internal factors to the brain.

					Monkeys		Humans
		Tested Range			M1	M2	
Model	Parameter	Min	Step	Max	Mean ± s.e.m.	Mean ± s.e.m.	Mean ± s.e.m.
VP-VP	$\bar{J}_{\text{high}}$	0	1.01	100	7.0 ± 1.9	6.8 ± 1.2	30.3 ± 3.6
	$\bar{J}_{\text{low}}$	0	1.01	100	1.08 ± 0.26	1.07 ± 0.24	1.72 ± 0.34
	τ	0.1	1.1	100	61 ± 18	63 ± 17	45 ± 11
	lapse	0	0.002	0.2	0.041 ± 0.038	0.044 ± 0.034	0.046 ± 0.018
VP-EP	$\bar{J}_{\text{high}}$	0	1.01	100	6.34 ± 0.92	9.2 ± 1.9	38.4 ± 4.4
	$\bar{J}_{\text{low}}$	0	1.01	100	1.09 ± 0.28	1.46 ± 0.54	2.32 ± 0.45
	τ	0.1	1.1	100	22.7 ± 4.7	41 ± 10	42.9 ± 7.8
	lapse	0	0.002	0.2	0.210 ± 0.002	0.19 ± 0.026	0.031 ± 0.013
VP-SP	$\bar{J}_{\text{high}}$	0	1.01	100	9.8 ± 3.7	7.85 ± 0.94	29.0 ± 3.1
	$\bar{J}_{\text{low}}$	0	1.01	100	4.1 ± 1.8	4.0 ± 0.47	5.35 ± 0.77
	τ	0.1	1.1	100	52 ± 25	70 ± 12	22.4 ± 5.4
	lapse	0	0.002	0.2	0.17 ± 0.061	0.011 ± 0.031	0.088 ± 0.018

Table 4.2. Parameter ranges and parameter estimates. For monkeys, means and standard errors were computed across 100 bootstrapped data sets. For humans, means and standard errors were computed across subjects.

These findings differ from those in a related change-detection task (Keshvari et al., 2012). Notably, the change-detection task used in the two studies differed in some important ways that changes the complexity of the task. First, the sample display in this study contained three items as opposed to four items in the Keshvari et al. (2012) study. Second, in the Keshvari study, subjects were asked to report whether or not a change occurred in two consecutive displays that were identical, except for one changed item in 50% of the trials. In the current study, there was always a change in the orientation of one of the items between displays and the task was to identify where the change occurred. Subjects in this study had a 50% chance of responding correctly, since only two items were shown during test. Third, in the Keshvari study, both the sample and test stimuli were presented for 100 ms each, whereas in this study, sample viewing time was 300 ms and the test stimuli remained on the screen until response. The latter two differences change the dependence of the task parameters and thus the generative model, which changes the inference and decision rules in many ways. Lastly, in the Keshvari (2012) task, subjects did not receive trial-to-trial feedback, whereas in this study, both humans (green light + tone) and monkeys (food reward) received trial-to-trial feedback. It has been suggested that when subjects receive trial-to-trial feedback to their responses, both humans and nonhuman animals could learn the values of the task variables, and use them strategically rather than relying on the internal estimates of uncertainty in the sensory measurements of those variables (Ma, 2012). For example, in our task, after receiving feedback, subjects (both humans and

monkeys) may have learned that they could maximize their performance by strategically allocating attention only to the high-reliability stimuli. It remains to be understood which of these differences could have contributed to the differences in the findings.

Change detection has primarily been used to understand the encoding limitations of VWM processing. This is the first account of testing monkeys in a change-detection task to understand whether monkeys use Bayesian-inference in perceptual decision-making. Our findings provide cross-species evidence that while humans and monkeys may be Bayesian observers, they may not always use probabilistic computations for optimal decision-making. In this regard, primates in general might be suboptimal, depending on the complexity of the task. It remains to be understood at which point optimality breaks down, even though probabilistic computations continue to be performed at the neural level.

CHAPTER 5: GENERAL CONCLUSIONS AND FUTURE DIRECTIONS

General Conclusions

The experiments presented in this thesis combined tools of psychophysics, learning, memory and computational modeling to compare for the first time encoding and decision processes of rhesus monkeys to those of humans in nearly identical visual working memory (VWM) tasks. In Aim 1 (Chapter 2), predictions of five leading models of VWM encoding limitations were tested in monkeys and humans using a change-detection task, where set size and change magnitude were systematically manipulated. In Aim 2 (Chapter 3), the change-detection task from Chapter 2 was modified to investigate whether monkeys and humans integrate uncertain sensory information from multiple items and whether they do so in an optimal fashion. Although change detection has been studied extensively in humans, several of these models had not been applied to the specific parameters of this task, and no previous study had compared all of these models in parallel with monkeys and humans.

I have shown here that in both species, resource models in general, provided a much better fit to the data than the classic item-limit ("slot") model. Resource models account for noisy encoding of stimuli, as opposed to an "all-or-none" storage. In accord with this theory, the errors in performance are then due to a problem in separating the signal from noise, rather than due to a fixed limit in the number of items that can be remembered. This notion fits well within the framework of signal detection theory and has been shown in many Bayesian models of perception, including change detection and its close variants (Elmore et al., 2011; Keshvari et al., 2012, 2013; Lara & Wallis, 2012; van den Berg, Shin,

et al., 2012; Wilken & Ma, 2004). A question that we do not formally address here is the source(s) of noise. Errors in recollecting stimuli from memory could be due to several factors that arise at different stages of memory processing; for example, noise during initial encoding of stimulus when sensory information is processed (particularly due to short viewing times), inability to maintain these memory representations when they are no longer accessible for view (delay period), difficulty retrieving memory of a sample stimulus, or deciding which response to make. We do not formally distinguish among these possibilities here; however, future studies might consider investigating these sources of noise, whether they can be manipulated or minimized, and how the nature of memory will be affected.

Specific to the class of resource models, results from both species support the notion of variability in memory precision, which allows for flexible allocation of precision for the encoding of most (if not all) items in the visual scene, as opposed to being equally distributed and/or with a fixed cap (Palmer, 1990; Zhang & Luck, 2008). We found that precision per item varies across stimuli and trials, and on average decreases with increasing set size. These findings are consistent with mounting evidence in humans across multiple tasks – change detection, change localization (color and orientation change), and delayed-estimation (color and orientation) -- providing further support for the variable precision point of view (Fougnie et al., 2012; Keshvari et al., 2012, 2013; van den Berg et al., 2014; van den Berg, Shin, et al., 2012). While the origins of the variability in precision are not completely understood, several factors could be at

play; for example, eye movements, fluctuations in attention over items and trials, differences in precision due to stimulus effects such as cardinal orientations and configural grouping, and variability in memory decay rates across stimuli (Brady & Tenenbaum, 2013; Fougnie et al., 2012; Girshick, Landy, & Simoncelli, 2011; Lara & Wallis, 2012; Nienborg & Cumming, 2009).

We considered two variants of the variable-precision model in Aim 1. One version was a variable-precision model with no fixed item limit, and the other was a hybrid model with an item limit. It is important to address the key parameters of these winning models. The parameter $\overline{J_1}$ that characterizes mean precision when set size is 1, differed greatly within individual subjects and between species. Differences in the values of mean precision could be attributed to several factors, including the stimulus-related effects, level of motivation, fatigue, or distraction. A second parameter, α, captures the effect of set size on mean precision, and describes the power with which precision per item decreases as set size increases. If set size had no effect on memory precision, then the value of α would be 0 (indeed, when set size was fixed in Aim 2 and thus had no effect on precision, we eliminated this parameter). We found that in both species and in both variants of the variable precision model, the α values were more negative than -1, indicating steep decreases in mean precision as set size increases.

In the variable-precision model with the fixed item limit, we found that *K* (typically taken to be the item limit number) was equal to about 3.5 in monkeys and 4.1 in humans. In initial reading, these values of *K* might be reminiscent of

the famous magical number 4 ± 1 items. However, it is important to realize that there is nothing magical about the values of this parameter. These values seem to vary largely across memory tasks, set sizes, as well as stimulus complexity, and are greatly dependent on which model is being considered; in fact, across monkeys and humans these values have been shown to vary anywhere from less than 1 to more than 6 items (Alvarez & Cavanagh, 2004; Buschman et al., 2011; Elmore et al., 2011; Eng et al., 2005; Heyselaar et al., 2011; Keshvari et al., 2013; Lara & Wallis, 2012; van den Berg et al., 2014; van den Berg, Shin, et al., 2012; Zhang & Luck, 2008). It is being suggested that if one were to adhere to the idea of a magical number for capacity, then the power value of α is a better replacement for *K*, as it provides a better characterization of the interplay between precision and set size to describe VWM limitations (Ma, 2014; under review). We do not interpret the value of K as the maximum limit on how many items can be remembered. Instead, we believe that it represents the number of items that an observer might process on a given trial (which could be a subset of total items presented), depending on their level of motivation, attention, or any strategic employments. Of course, memory resource (and precision) is limited, and as the amount of information presented increases, this resource may not in practice be infinitely divisible. In such cases, it is possible that observers flexibly allocate memory resource to a subset of visual stimuli (which could be an item or a combination of stimulus features as when viewing natural scenes).

In Aim 2, we found that monkeys and humans are Bayesian-observers and take into account the uncertainty of sensory observations when making

perceptual judgments, giving more weight to more certain evidence. However, while memory precision varies on an item-by-item and trial-by-trial basis, in this task, humans and monkeys seemed to make an incorrect assumption about precision, and assume two values of precision that are completely determined by the reliability of the stimulus. They ignore any additional variability in precision that may arise due to noise from both external and internal factors to the brain. This begs the question, how can observers make a wrong (or right) assumption about precision? In probabilistic models, it is suggested that neural populations encode with probability distributions over stimulus values on a trial-by-trial and item-by-item basis (rather than using point estimates). This "implicit" knowledge of the internal representations of stimuli can then be used in downstream computation for perceptual judgment and decision-making in optimal or suboptimal ways. Since these computations have been shown to differ based on the complexity of the tasks in both humans and monkeys, it is important to investigate which factors determine whether optimal or suboptimal decision rules are used (Gu, Angelaki, & DeAngelis, 2008; Keshvari et al., 2012; Ma & Jazayeri, 2014; van den Berg, Vogel, Josic, & Ma, 2012; Yang & Shadlen, 2007).

We think that the most remarkable findings in these series of experiments are the qualitative similarities that these two primate species share in both encoding and decision-making. In both aims, we found that despite the quantitative performance differences between the two species, the same winning models account for the species' behavior. The qualitative similarities between the two species shown here may have been expected because rhesus monkeys

have been shown to share many aspects of memory processes with humans. For example, monkeys show the same primary and recency effects in serial position functions and show striking similarities with humans providing evidence for a continuous-resource model (Elmore et al., 2011; Roberts & Kraemer, 1981; Sands & Wright, 1980; Wright, Santiago, Sands, Kendrick, & Cook, 1985). Nevertheless, it is one thing to expect similarities from our experiments, but it is another thing altogether to show them and show just how extensive they are.

These earlier findings along with ours are suggestive of evolutionary continuity and common underlying mechanisms of VWM processes in primates generally. Indeed, rhesus monkeys and humans have similar neural architecture, especially the visual cortex and areas of the prefrontal cortex that are relevant to VWM processing (Funahashi et al., 1989; Orban, Van Essen, & Vanduffel, 2004; Petrides, 1996). Our results, thus, establish rhesus monkeys as a suitable model system for further elucidating the neural substrates of VWM limitations in encoding and decision-making. The combination of psychophysics, computational modeling, and neurophysiological methods offers great potential to unravel how VWM works and why it fails.

Future directions

It is encouraged that future studies using change detection not only consider the typical manipulation of set size, but also vary change magnitude. As demonstrated here, the addition of the change-magnitude manipulation unveils

profound differences in model predictions. In the same vein, future studies aimed at understanding the mechanisms of VWM should provide direct comparisons across many if not all leading models (than considering just one or two), because only then can goodness-of-fit be properly compared across models and arguments of specific processing. Indeed, a recent study that conducted a factorial comparison of 32 models that reanalyzed data from several studies of VWM limitations found that the conclusions of these data could greatly differ from previously-made claims, when such an approach is used (van den Berg et al., 2014).

The studies described here (and a growing body of research in humans) have pointed to the variability in precision as the key factor in characterizing VWM limitations, although the neural basis of variability in precision is only beginning to be explored (Emrich, Riggall, Larocque, & Postle, 2013; Ester, Anderson, Serences, & Awh, 2013). In general, the neural substrates of VWM limitations have previously been based on the framework of fixed-capacity models. But now with a growing body of evidence in humans for the variable-precision model, coupled with identical findings in monkeys from this study, there is convergent evidence for the fundamental underlying processes of VWM that will serve as a standard for guiding and interpreting neurobiological investigations of the complex maze of neural circuits responsible for memory.

BIBLIOGRAPHY

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. doi:10.1109/TAC.1974.1100705

Alescio-Lautier, B., Michel, B. F., Herrera, C., Elahmadi, A., Chambon, C., Touzet, C., & Paban, V. (2007). Visual and visuospatial short-term memory in mild cognitive impairment and Alzheimer disease: role of attention. *Neuropsychologia*, *45*(8), 1948–1960. doi:10.1016/j.neuropsychologia.2006.04.033

Alvarez, G. A., & Cavanagh, P. (2004). 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.

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology*, *19*(4), 452–458. doi:10.1016/j.conb.2009.06.008

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628. doi:10.1111/j.1467-9280.2007.01949.x

Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*(1), 201–208. doi:10.1016/j.neuroscience.2005.08.023

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. doi:10.1126/science.1736359

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *The Journal of Neuroscience*, *34*(10), 3632–3645. doi:10.1523/JNEUROSCI.3204-13.2014

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7–7. doi:10.1167/9.10.7

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science (New York, N.Y.)*, *321*(5890), 851–854. doi:10.1126/science.1158023

Berryhill, M. E., & Olson, I. R. (2008). The right parietal lobe is critical for visual working memory. *Neuropsychologia*, *46*(7), 1767–1774. doi:10.1016/j.neuropsychologia.2008.01.009

Brady, T. F., & Tenenbaum, J. B. (2013). 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

Brouwer, A. M., & Knill, D. C. (2007). The role of memory in visually guided reaching. *Journal of Vision*, *7*(5), 6–6. doi:10.1167/7.5.6

Burnham, K. P. (2002). *Model selection and multimodel inference: a practical information-theoretic approach* (2nd ed.). New York: Springer.

Buschman, T. J., Siegel, M., Roy, J. E., & Miller, E. K. (2011). Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of*

Sciences of the United States of America, *108*(27), 11252–11255. doi:10.1073/pnas.1104666108

Christopher, G., & MacDonald, J. (2005). The impact of clinical depression on working memory. *Cognitive Neuropsychiatry*, *10*(5), 379–399. doi:10.1080/13546800444000128

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia*, *49*(6), 1407–1409. doi:10.1016/j.neuropsychologia.2011.01.029

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology. Human Perception and Performance*, *21*(1), 109–127.

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*(6), 693–702. doi:10.1038/nn.2123

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). 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). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. doi:10.1017/S0140525X01003922

Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.

Cowan, N. (2011). The focus of attention as observed in visual working memory tasks: making sense of competing claims. *Neuropsychologia*, *49*(6), 1401–1406. doi:10.1016/j.neuropsychologia.2011.01.035

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. doi:10.1016/j.cogpsych.2004.12.001

Efron, B. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Elmore, L. C., Ma, W. J., Magnotti, J. F., Leising, K. J., Passaro, A. D., Katz, J. S., & Wright, A. A. (2011). 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, S. M., Riggall, A. C., Larocque, J. J., & Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(15), 6516–6523. doi:10.1523/JNEUROSCI.5732-12.2013

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, *12*(6), 1127–1133.

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

Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A neural measure of precision in visual working memory. *Journal of Cognitive Neuroscience*, *25*(5), 754–761. doi:10.1162/jocn_a_00357

Ezzyat, Y., & Olson, I. R. (2008). The medial temporal lobe and visual working memory: Comparisons across tasks, delays, and visual similarity. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(1), 32–40. doi:10.3758/CABN.8.1.32

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, *9*(4), 292–303. doi:10.1038/nrn2258

Farmer, C. M., O'Donnell, B. F., Niznikiewicz, M. A., Voglmaier, M. M., McCarley, R. W., & Shenton, M. E. (2000). Visual perception and working memory in schizotypal personality disorder. *The American Journal of Psychiatry*, *157*(5), 781–788.

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229. doi:10.1038/ncomms2237

Fukuda, K., & Vogel, E. K. (2011). Visual short term memory also gates long term memory without explicit retrieval. *Journal of Vision*, *11*(11), 1275–1275. doi:10.1167/11.11.1275

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61*(2), 331–349.

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

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932. doi:10.1038/nn.2831

Gold, J. M., Wilk, C. M., McMahon, R. P., Buchanan, R. W., & Luck, S. J. (2003). Working memory for visual features and conjunctions in schizophrenia. *Journal of Abnormal Psychology*, *112*(1), 61–71.

Goldman-Rakic, P. S. (1990). Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates. *Progress in Brain Research*, *85*, 325–335; discussion 335–336.

Green, J.M. and Swets, D.M. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons Inc.

Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, *11*(10), 1201–1210. doi:10.1038/nn.2191

Henderson, J. (2008). Eye movements and visual memory. In *Visual memory* (S. Luck & A. Hollingworth (Eds.)., pp. 87–121). Oxford: Oxford University Press.

Heyselaar, E., Johnston, K., & Paré, M. (2011). A change detection approach to study visual working memory of the macaque monkey. *Journal of Vision*, *11*(3). doi:10.1167/11.3.11

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307. doi:10.1093/biomet/76.2.297

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*(3), 420–456.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304. doi:10.1146/annurev.psych.55.090902.142005

Keshvari, S., van den Berg, R., & Ma, W. J. (2012). Probabilistic computation in human perception under variability in encoding precision. *PLoS ONE*, *7*(6), e40216. doi:10.1371/journal.pone.0040216

Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No evidence for an item limit in change detection. *PLoS Computational Biology*, *9*(2), e1002927. doi:10.1371/journal.pcbi.1002927

Kim, S., Liu, Z., Glizer, D., Tannock, R., & Woltering, S. (2014). Adult ADHD and working memory: neural evidence of impaired encoding. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *125*(8), 1596–1603. doi:10.1016/j.clinph.2013.12.094

Kiyonaga, A., & Egner, T. (2014). The Working Memory Stroop Effect: When Internal Representations Clash With External Stimuli. *Psychological Science*, *25*(8), 1619–1629. doi:10.1177/0956797614536739

Knill, D. C., & Pouget, A. (2004). 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

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge, U.K. ; New York: Cambridge University Press.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. doi:10.1038/nature02169

Lara, A. H., & Wallis, J. D. (2012). Capacity and precision in an animal model of visual short-term memory. *Journal of Vision*, *12*(3). doi:10.1167/12.3.13

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. doi:10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. doi:10.1016/j.tics.2013.06.006

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge, UK ; New York: Cambridge University Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: a user's guide* (2nd ed.). Mahwah, N.J: Lawrence Erlbaum Associates.

Ma, W. J. (2012). Organizing probabilistic models of perception. *Trends in Cognitive Sciences*, *16*(10), 511–518. doi:10.1016/j.tics.2012.08.010

Ma, W. J. (2014).The magical number -1.3, plus or minus 0.6: new limits on our capacity for processing information. under review.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. doi:10.1038/nn.3655

Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual Review of Neuroscience*, *37*, 205–220. doi:10.1146/annurev-neuro-071013-014017

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *The Journal of Neuroscience*, *16*(16), 5154–5167.

Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Nienborg, H., & Cumming, B. G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature*, *459*(7243), 89–92. doi:10.1038/nature07821

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception & Psychophysics*, *64*(7), 1055–1067.

Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences*, *8*(7), 315–324. doi:10.1016/j.tics.2004.05.009

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(2), 332–350.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378.

Pellicano, E., Gibson, L., Maybery, M., Durkin, K., & Badcock, D. R. (2005). Abnormal global processing along the dorsal visual pathway in autism: a possible mechanism for weak visuospatial coherence? *Neuropsychologia*, *43*(7), 1044–1053. doi:10.1016/j.neuropsychologia.2004.10.003

Petrides, M. (1996). Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *351*(1346), 1455–1461; discussion 1461–1462. doi:10.1098/rstb.1996.0130

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2), 283–290.

Pisella, L., Berberovic, N., & Mattingley, J. B. (2004). Impaired working memory for location but not for colour or shape in visual neglect: a comparison of parietal and non-parietal lesions. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *40*(2), 379–390.

Rainer, G., Asaad, W. F., & Miller, E. K. (1998). Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature*, *393*(6685), 577–579. doi:10.1038/31235

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277. doi:10.1146/annurev.psych.53.100901.135125

Roberts, W. A., & Kraemer, P. J. (1981). Recognition memory for lists of visual stimuli in monkeys and humans. *Animal Learning & Behavior*, *9*(4), 587–594.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(16), 5975–5979. doi:10.1073/pnas.0711295105

Sands, S. F., & Wright, A. A. (1980). Serial probe recognition performance by a rhesus monkey and a human with 10- and 20-item lists. *Journal of Experimental Psychology. Animal Behavior Processes*, *6*(4), 386–396.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. doi:10.1214/aos/1176344136

Shaw, M. (1980). Identifying attentional and decision-making components in information processing. In *Attention and Performance* (pp. 277–296). Hillsdale, NJ: Erlbaum.

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

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, *23*(8), 775–785.

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

Van den Berg, R., & Ma, W. J. (2014). “Plateau”-related summary statistics are uninformative for comparing working memory models. *Attention, Perception, & Psychophysics*. doi:10.3758/s13414-013-0618-7

Van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785. doi:10.1073/pnas.1117465109

Van den Berg, R., Vogel, M., Josic, K., & Ma, W. J. (2012). Optimal inference of sameness. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(8), 3178–3183. doi:10.1073/pnas.1108790109

Vasterling, J. J., Brailey, K., Constans, J. I., & Sutker, P. B. (1998). Attention and memory dysfunction in posttraumatic stress disorder. *Neuropsychology*, *12*(1), 125–133.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751. doi:10.1038/nature02447

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of*

Experimental Psychology. Human Perception and Performance, *27*(1), 92–114.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135. doi:10:1167/4.12.11

Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., & Cook, R. G. (1985). Memory processing of serial lists by pigeons, monkeys, and people. *Science (New York, N.Y.)*, *229*(4710), 287–289.

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

Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*(7148), 1075–1080. doi:10.1038/nature05852

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. doi:10.1038/nature06860

VITA

Deepna Devkar was born in Gujarat, India on September 2, 1985, the daughter of Tribhovan and Rekha Thakkar. She moved to Houston, Texas in 1999. After graduating from Westfield High School in 2002, she attended Baylor University, in Waco, Texas. She received a Bachelor of Science with a major in Psychology in December 2006. For the following year, she worked as a research coordinator at Baylor College of Medicine. She received her Masters degree in Psychology from the University of Texas, San Antonio. In August of 2010, she entered the Graduate School of Biomedical Sciences at The University of Texas Health Science Center at Houston (UTHSC) to begin work on her doctorate in Neuroscience. Upon completion of her Ph.D., she plans to accept a post-doctoral research fellowship at New York University. .

Permanent address:

403 Orchard Ln,

Sugar Land, TX 77498