

**MEASURING AND MODELING OF HUMAN VISUAL SHORT-
TERM MEMORY**

A Dissertation Submitted to the Faculty of

The Graduate School
Baylor College of Medicine

In Partial Fulfillment of the
Requirements for the Degree

of

Doctor of Philosophy

by

HONGSUP SHIN

Houston, Texas

April 20, 2015

APPROVED BY THE DISSERTATION COMMITTEE

Signed _____

Wei Ji Ma
Chairman

Dora Angelaki

Fabrizio Gabbiani

Andreas Tolias

Michael Beauchamp

APPROVED BY THE DEPARTMENT OF NEUROSCIENCE

Signed _____

Paul Pffafinger
Director of Graduate Studies

Dora Angelaki
Chairman of Department

APPROVED BY THE DEAN OF GRADUATE SCIENCES

Signed _____

Deborah L. Johnson, Ph.D.

Date _____

Acknowledgments

I would like to thank my supervisor, Wei Ji Ma for his guidance and feedback as an exceptional mentor. I would also like to thank Dora Angelaki, Michael Beauchamp, Fabrizio Gabbiani, and Andreas Tolias for their intellectual input and valuable opinions as my thesis committee members.

I would like to thank Ronald van den Berg for helping me learn programming, design experiments, and for being a good mentor and a colleague. I am also grateful to have great lab colleagues in Baylor College of Medicine and New York University. I would also like to express my gratitude to my fellow graduate students at Baylor College of Medicine, especially for their participation in my experiments. I appreciate all the help I received from the administrative staff at Baylor College of Medicine especially for dealing with on-site issues while I was a remote student.

Finally, I could not have gone through many journeys during this research without the support of my family and friends. I would also like to thank my partner, Gabriel Bodeen for his patience and endless support.

Abstract

Visual short-term memory (VSTM) is a temporary buffer for storing visual information. VSTM is impaired in conditions such as schizophrenia and change blindness. However, its basic computational mechanisms are still insufficiently understood. Many psychology and cognitive neuroscience studies have assumed that the information is stored in discrete, fixed number of slots. This idea has affected many aspects of VSTM studies, such as those studying the effect of the number of items (set size), the number of features (e.g. color and orientation), and maintenance duration (delay time). However, recent studies have suggested that VSTM is noisy and is better conceptualized as a continuous resource that is distributed to all items.

In my work, I developed a new quantitative model of VSTM, the variable-precision (VP) model, in which encoding precision is a continuous quantity that varies among trials and items and on average decreases with increasing set size. By using a change localization task and performing formal model comparison, I found that the VP model outperforms other classic VSTM models in accounting for human data. In order to more exhaustively explore model space, I developed hybrid models by combining components of existing basic models. I found models that have variability either in precision or in the number of encoded items can better explain change localization behavior than models that do not.

Since natural objects have more than one feature, I have extended my work to orientation-color stimuli. For objects with multiple features, specifically orientation and color, I reformulated old questions based on the “slot” idea as new hypotheses based on the idea that memories are noisy. My results suggest that orientation and color have independent pools of resource, but that some portion of each pool is packaged together and divided among all objects, while the rest is allocated only to objects in which the corresponding feature is relevant.

Finally, by using a delayed-estimation task, I examined the effect of delay duration on VSTM decay. The results show that estimation error gradually increases with delay duration and set size.

Taken together, my results lay the groundwork for an understanding of VSTM rooted in terms of the precision of encoding.

Table of Contents

Approvals.....	2
Acknowledgments	3
Abstract.....	4
Table of Contents	6
List of Figures.....	12
List of Tables	19
Chapter 1 General introduction	21
1 History of research on VSTM.....	22
2 History of main paradigms	23
2.1 Change detection.....	24
2.2 Delayed estimation.....	25
2.3 Change localization.....	25
3 History of VSTM models.....	28
3.1 Discrete-capacity models.....	28
3.2 Continuous-resource models	29
4 Overview of chapters	32
Chapter 2 Modeling set-size effect of VSTM.....	33

1	Introduction.....	34
2	Summary of task	34
3	Theory and models.....	35
3.1	Step 1: Encoding: Measurement	35
3.2	Step 1: Encoding: Precision and resource	38
3.3	Step 1: Encoding: Models of memory	39
3.4	Step 2: Inference	44
3.5	Step 3: Response probabilities.....	47
3.6	Item-limit model	49
4	Methods.....	50
4.1	Psychophysics.....	50
4.2	Model comparison	51
5	Results	53
5.1	Summary statistics.....	53
5.2	Model comparison	58
5.3	Apparent guessing rate	62
6	Discussion.....	64
Chapter 3 Factorial comparison of VSTM models.....		65
1	Introduction.....	66
2	Summary of tasks.....	67
3	Theory and models.....	67
3.1	Model factors.....	67

3.2	Factorial combination of the modeling dimensions.....	69
3.3	Nomenclature	71
3.4	Models.....	72
3.5	Bayesian inference in change detection	75
4	Methods.....	77
4.1	Psychophysics.....	77
4.2	Model comparison	77
4.3	Model recovery analysis.....	79
5	Results	81
5.1	Model comparison	81
5.2	Statistical dependence across tasks and features.....	92
6	Discussion.....	93
Chapter 4 VSTM for two-feature objects (1): memory allocation to relevant and irrelevant locations.....		
95		
1	Introduction.....	96
2	Theory and models.....	99
2.1	From resource to encoding precision.....	100
2.2	Step 1: Encoding.....	101
2.3	Step 2: Inference	103
2.4	Step 3: Response probabilities.....	104
3	Methods.....	107
3.1	Laboratory experiments.....	107

3.2	Amazon Mechanical Turk experiments	112
4	Results	114
4.1	Answering Q1: Resource is not shared between orientation and color.	114
4.2	Answering Q2: The irrelevant feature does not affect performance.	119
4.3	Answering Q2's follow-up question: The irrelevant feature is encoded but ignored.	124
5	Discussion.....	128
Chapter 5 VSTM for two-feature objects (2): spatial allocation of memory resource		130
1	Introduction.....	131
2	Theory and models.....	132
2.1	No-leak model	132
2.2	Leak model	133
3	Methods.....	133
3.1	Psychophysics.....	133
3.2	Experimental conditions	134
4	Results	136
4.1	Summary statistics.....	136
4.2	Model comparison	139
5	Discussion.....	141
Chapter 6 The temporal decay of VSTM		142
1	Introduction.....	143

2	Summary of task	144
3	Theory and models.....	144
3.1	Variable-precision model.....	144
4	Methods.....	145
4.1	Psychophysics.....	145
5	Results	145
5.1	Estimate error distribution.....	145
5.2	Maximum-likelihood estimates of model parameters	150
6	Discussion.....	153
Chapter 7 Summary and conclusion		154
1	Variable precision of visual short-term memory	155
1.1	Do slots exist?	155
1.2	Is resource discrete?	156
1.3	Exploration of further model dimensions	156
1.4	Neural basis of VSTM resource	158
1.5	Decrease of mean precision with set size	159
2	Visual short-term memory for more complex objects.....	159
2.1	New framework for VSTM of orientation and color.....	162
2.2	VSTM for objects with other feature combinations.....	163
2.3	Definition of an object	166
2.4	Location as a feature	167
2.5	Color as an ecological relevant feature.....	167

3	Temporal dynamics of VSTM.....	168
4	Conclusion	169
	Bibliography	171

List of Figures

FIGURE 1.1 TRIAL PROCEDURE OF CHANGE LOCALIZATION TASK. (A) COLOR CHANGE LOCALIZATION. (B) ORIENTATION CHANGE LOCALIZATION.....27

FIGURE 1.2 MODELS OF VSTM (IL: THE ITEM-LIMIT MODEL, SA: THE SLOTS-PLUS-AVERAGING MODEL, EP: THE EQUAL-PRECISION MODEL, VP: THE VARIABLE-PRECISION MODEL). EACH BOX REPRESENTS AN ITEM. CAPACITY OR RESOURCE IS REPRESENTED GRAY. SET SIZE IS 2 (LEFT) OR 5 (RIGHT). IN THIS EXAMPLE, THE NUMBER OF “SLOTS” OR “CHUNKS” IS 3 IN THE IL AND SA MODELS.31

FIGURE 2.1 THEORY. (A) TOP: IN THE VP MODEL, PRECISION, J , IS VARIABLE AND ASSUMED TO FOLLOW A GAMMA DISTRIBUTION (HERE WITH $\tau=1$). BOTTOM: VON MISES NOISE DISTRIBUTIONS CORRESPONDING TO THREE VALUES OF PRECISION AND $s=0$. A HIGHER J PRODUCES A NARROWER DISTRIBUTION $P(x|s,J)$. (B) EXAMPLE PROBABILITY DISTRIBUTIONS OVER PRECISION AT DIFFERENT SET SIZES IN THE VP MODEL. HERE, MEAN PRECISION (DASHED LINES) WAS TAKEN INVERSELY PROPORTIONAL TO SET SIZE ($\alpha=1$). IN THE EP MODEL, THESE DISTRIBUTIONS WOULD BE DELTA FUNCTIONS.43

FIGURE 2.2 GENERATIVE MODEL FOR THE CHANGE DETECTION TASK. C : CHANGE OCCURRENCE (0: NO CHANGE; 1: CHANGE) ; Δ : MAGNITUDE OF CHANGE; $\vec{\Delta}$: VECTOR OF

CHANGE MAGNITUDES AT ALL LOCATIONS; θ AND ϕ : VECTORS OF STIMULI IN THE FIRST AND SECOND DISPLAYS, RESPECTIVELY; X AND Y: VECTORS OF MEASUREMENTS IN THE FIRST AND SECOND DISPLAYS, RESPECTIVELY46

FIGURE 2.3 PROPORTION CORRECT (SUBJECT AVERAGE) AS A FUNCTION OF CHANGE MAGNITUDE AT EACH SET SIZE IN THE COLOR AND ORIENTATION EXPERIMENTS....54

FIGURE 2.4 INDIVIDUAL-SUBJECT FITS OF THE SA AND VP MODELS IN THE COLOR EXPERIMENT (SOLID: VP; DASHED: SA; OTHER MODELS ARE NOT SHOWN TO AVOID CLUTTER). (A) PROPORTION CORRECT AS A FUNCTION OF SET SIZE. (B) PROPORTION CORRECT AS A FUNCTION OF CHANGE MAGNITUDE AT EACH SET SIZE.55

FIGURE 2.5 INDIVIDUAL-SUBJECT FITS OF THE SA AND VP MODELS IN THE ORIENTATION EXPERIMENT (SOLID: VP; DASHED: SA; OTHER MODELS ARE NOT SHOWN TO AVOID CLUTTER). (A) PROPORTION CORRECT AS A FUNCTION OF SET SIZE. (B) PROPORTION CORRECT AS A FUNCTION OF CHANGE MAGNITUDE AT EACH SET SIZE.56

FIGURE 2.6 MODEL LOG LIKELIHOODS RELATIVE TO THE VP MODEL. (A) COLOR EXPERIMENT. (B) ORIENTATION EXPERIMENT.59

FIGURE 2.7 BAYESIAN MODEL COMPARISON RESULTS FOR INDIVIDUAL SUBJECTS IN DELAYED ESTIMATION. (A) COLOR EXPERIMENT. (B) ORIENTATION EXPERIMENT.60

FIGURE 2.8 BAYESIAN INFORMATION CRITERION FOR INDIVIDUAL SUBJECTS IN DELAYED ESTIMATION. (A) COLOR EXPERIMENT. (B) ORIENTATION EXPERIMENT.....61

FIGURE 2.9 APPARENT GUESSING RATE AS A FUNCTION OF SET SIZE. (A) COLOR EXPERIMENT. (B) ORIENTATION EXPERIMENT.63

FIGURE 3.1 MODELS OF VSTM. BASED ON TWO MODELING CRITERIA (ENCODING PRECISION AND NUMBER OF ITEMS ENCODED), 14 COMBINATION OF MODELS (CIRCLES) ARE SUGGESTED FOR EXHAUSTIVE MODEL COMPARISON. BASIC VSTM MODELS FROM CHAPTER 2 ARE IPF, SAF, EPA AND VPA.....73

FIGURE 3.2 GENERATIVE MODEL FOR THE CHANGE DETECTION TASK. C : CHANGE OCCURRENCE (0: NO CHANGE; 1: CHANGE); Δ : MAGNITUDE OF CHANGE; $\vec{\Delta}$: VECTOR OF CHANGE MAGNITUDES AT ALL LOCATIONS; θ AND ϕ : VECTORS OF STIMULI IN THE FIRST AND SECOND DISPLAYS, RESPECTIVELY; x AND y : VECTORS OF MEASUREMENTS IN THE FIRST AND SECOND DISPLAYS, RESPECTIVELY.....76

FIGURE 3.3 TRIAL PROCEDURE OF CHANGE DETECTION. (A) COLOR EXPERIMENT. (B) ORIENTATION EXPERIMENT.78

FIGURE 3.4 MODEL RECOVERY ANALYSIS. WE TESTED HOW WELL SYNTHETIC DATA SETS GENERATED FROM EACH MODEL (COLUMNS) WERE FITTED BY EACH MODEL (ROWS). THE COLOR IN A CELL INDICATES LOG LIKELIHOOD (LOG BAYES FACTOR) OF ALL FITTED MODELS RELATIVE TO THE WINNING MODEL FOR THE CORRESPONDING SYNTHETIC DATA. THE DARKEST RED ON THE DIAGONAL MEANS THAT THE MODEL USED TO GENERATE THE DATA WAS FOUND TO BE MOST LIKELY.....80

FIGURE 3.5 RELATIVE LOG LIKELIHOOD OF MODELS TO THE WINNING MODEL IN DESCENDING ORDER FOR THE ORIENTATION (THE SAP) AND COLOR (THE VPP) CHANGE LOCALIZATION EXPERIMENTS. MODELS WITH BOLD LETTERS ARE THE ONES USED IN CHAPTER 2.83

FIGURE 3.6 MODEL FITS TO SUBJECT PERFORMANCE FROM THE ORIENTATION CHANGE LOCALIZATION EXPERIMENT. THE NAMES OF THE MODELS TESTED IN CHAPTER ARE MARKED WITH *. THE ROOT-MEAN-SQUARE ERROR (RMSE) IS SHOWN FOR EACH MODEL.....	85
FIGURE 3.7 MODEL FITS TO SUBJECT PERFORMANCE FROM THE COLOR CHANGE LOCALIZATION EXPERIMENT.....	86
FIGURE 3.8 RELATIVE LOG LIKELIHOOD OF MODELS TO THE WINNING MODEL (THE VP-F) IN DESCENDING ORDER FOR THE ORIENTATION AND COLOR CHANGE DETECTION EXPERIMENTS. MODELS WITH BOLD LETTERS ARE THE ONES USED IN CHAPTER 2.	88
FIGURE 3.9 MODEL FITS TO SUBJECT PERFORMANCE FROM THE ORIENTATION CHANGE DETECTION EXPERIMENT.....	90
FIGURE 3.10 MODEL FITS TO SUBJECT PERFORMANCE FROM THE COLOR CHANGE DETECTION EXPERIMENT.....	91
FIGURE 4.1 MODELS OF VSTM FOR Q1 AND Q2 AND THEIR PREDICTIONS FOR PERFORMANCE IN DIFFERENT CONDITIONS.....	106
FIGURE 4.2 TRIAL PROCEDURES IN EXPERIMENTS 1, 2, AND 3. SUBJECTS CLICK ON THE LOCATION WHERE A (RELEVANT) CHANGE HAS OCCURRED. (A) ORIENTATION TRIAL OF CONDITION A: ORIENTATION IS THE ONLY FEATURE. (B) CONDITION B: STIMULI HAVE TWO FEATURES AND THE CHANGE HAS OCCURRED IN EITHER. (C) ORIENTATION TRIAL OF CONDITION C: STIMULI HAVE TWO FEATURES BUT ONLY ORIENTATION IS RELEVANT. THE IRRELEVANT FEATURE (HERE COLOR) DOES NOT CHANGE. (D) ORIENTATION TRIAL	

OF CONDITION D: AS C, BUT THE IRRELEVANT FEATURE VALUE (HERE COLOR) HAS CHANGED IN EITHER ONE OR ALL STIMULI.	110
FIGURE 4.3 EXPERIMENT 1. PROPORTION CORRECT AS A FUNCTION OF CHANGE MAGNITUDE (8 SUBJECTS), WITH MODEL FITS (SHADED AREAS: MEAN \pm S.E.M.). PERFORMANCE WAS INDISTINGUISHABLE BETWEEN CONDITIONS B AND C, IN ACCORDANCE WITH MODEL 5/6 BUT NOT WITH MODEL 3.	116
FIGURE 4.4 EXPERIMENT 2. PROPORTION CORRECT AS A FUNCTION OF CHANGE MAGNITUDE (8 SUBJECTS), WITH MODEL FITS (SHADED AREAS: MEAN \pm S.E.M.). PERFORMANCE WAS INDISTINGUISHABLE BETWEEN CONDITIONS A AND C, IN ACCORDANCE WITH MODEL 5/6 BUT NOT WITH MODELS 1 AND 2.	118
FIGURE 4.5 PROPORTION CORRECT AS A FUNCTION OF CHANGE MAGNITUDE (5 SUBJECTS), WITH MODEL FITS (SHADED AREAS: MEAN \pm S.E.M.). PERFORMANCE WAS INDISTINGUISHABLE BETWEEN CONDITIONS C AND D, IN ACCORDANCE WITH MODEL 5/6 BUT NOT WITH MODEL 4.	121
FIGURE 4.6 DELAYED ESTIMATION TASK. SUBJECTS PROBE AN ESTIMATE OF THE RELEVANT FEATURE FOR THE FIRST 30 TRIALS, AND ON THE SURPRISE 31 ST TRIAL, THEY ARE ASKED TO PROBE THE IRRELEVANT FEATURE.	126
FIGURE 4.7 EXPERIMENT 4. THE ERROR DISTRIBUTION FROM POOLED DATA (600 SUBJECTS RESPECTIVELY) IS SHOWN BELOW THE PROCEDURE. (A) RELEVANT FEATURE: ORIENTATION, IRRELEVANT FEATURE: COLOR. BOTH DISTRIBUTIONS ARE SIGNIFICANTLY DIFFERENT FROM UNIFORM. (B) RELEVANT FEATURE: COLOR,	

IRRELEVANT FEATURE: ORIENTATION. BOTH DISTRIBUTIONS ARE SIGNIFICANTLY DIFFERENT FROM UNIFORM.	127
FIGURE 5.1 TRIAL PROCEDURES. A CHANGE OCCURRED EITHER IN ORIENTATION OR COLOR. SUBJECTS CLICK ON THE LOCATION WHERE THE CHANGE OCCURRED.	135
FIGURE 5.2 PSYCHOMETRIC CURVES AND MODEL PREDICTIONS. THE NO-LEAK MODEL (TOP) FITS BETTER THAN THE LEAK MODEL (BOTTOM).	137
FIGURE 6.1 HISTOGRAMS OF ERRORS OF AN INDIVIDUAL SUBJECT; SET SIZE (COLUMNS) AND DURATION OF DELAY (ROWS). CIRCULAR STANDARD DISTRIBUTION IN RADIANS IS SHOWN FOR EACH CONDITION. THE RED LINE SHOWS THE MODEL FITS OF THE VARIABLE-PRECISION (VP) MODEL.	147
FIGURE 6.2 AVERAGE CSD ACROSS SUBJECTS FOR EACH SET SIZE AND DURATION OF DELAY IN RADIANS.	148
FIGURE 6.3 (A) MEAN AND S.E.M. OF MAXIMUM-LIKELIHOOD ESTIMATES (MLEs) OF PARAMETERS OF THE VP MODEL. LEFT: MEAN PRECISION (\bar{J}_1), MIDDLE: POWER PARAMETER (A), RIGHT: SCALE PARAMETER (T). (B) MEAN AND S.E.M. OF (MLEs) OF PARAMETERS OF THE MIXTURE MODEL: PRECISION (J_1) AND THE MIXTURE PROPORTION OF THE VON MISES COMPONENT (w). (C) COEFFICIENT OF VARIATION (CV) OF A GAMMA DISTRIBUTION FROM THE VP MODEL ($CV = \sqrt{\frac{\tau}{J_1}}$).....	152
FIGURE 7.1 NEW FRAMEWORK FOR VSTM OF ORIENTATION AND COLOR. (A) FEATURE RESOURCE CONSISTS OF “PACKAGED” RESOURCE, WHICH GETS ALLOCATED TO ALL OBJECTS, AND “TARGETED” RESOURCE, WHICH ONLY GETS ALLOCATED TO OBJECTS	

FOR WHICH THAT FEATURE IS RELEVANT. (B) APPLICATION TO THE CONDITIONS FROM
CHAPTER 4.165

List of Tables

TABLE 2.1 MEANS AND SEs OF THE MAXIMUM-LIKELIHOOD ESTIMATES AND TESTED RANGES OF MODEL PARAMETERS.	57
TABLE 3.1 SUMMARY OF THE VSTM MODELS TESTED IN CHAPTER 3.	74
TABLE 3.2 MEAN AND S.E.M. OF RELATIVE LOG LIKELIHOOD OF EACH MODEL TO THE WINNING MODEL FOR THE ORIENTATION (THE SAP) AND COLOR (THE VPP) CHANGE LOCALIZATION EXPERIMENTS. MODELS WITH BOLD LETTERS GIVE INDISTINGUISHABLE MODEL PREDICTION COMPARED TO THE WINNING MODEL. MODELS THAT WERE USED IN CHAPTER 2 ARE MARKED WITH ASTERISKS.	84
TABLE 3.3 MEAN AND S.E.M. OF RELATIVE LOG LIKELIHOOD OF EACH MODEL TO THE WINNING MODEL (THE VP-F) FOR THE ORIENTATION AND COLOR CHANGE DETECTION EXPERIMENTS. MODELS WITH BOLD LETTERS GIVE INDISTINGUISHABLE MODEL PREDICTION COMPARED TO THE WINNING MODEL. MODELS THAT WERE USED IN CHAPTER 2 ARE MARKED WITH ASTERISKS.	89
TABLE 4.1 MEAN AND S.E.M. OF MAXIMUM-LIKELIHOOD ESTIMATES OF PARAMETERS.	122
TABLE 4.2 INDIVIDUAL AIC VALUES OF ALL MODELS FROM EXPERIMENTS 1, 2, AND 3.	123
Table 5.1 Means and SEs of the maximum-likelihood estimates of model parameters.	

TABLE 6.1 MEAN AND S.E.M. OF CIRCULAR STANDARD DEVIATION ACROSS SUBJECTS FROM EACH SET SIZE AND DURATION OF DELAY COMBINATION	149
TABLE 6.2 MEAN AND S.E.M. OF MAXIMUM-LIKELIHOOD ESTIMATES OF THE PARAMETERS IN THE VP MODEL.	151
TABLE 7.1 COMPARISON OF THE PREVIOUSLY POSED HYPOTHESES (H1-H3) ON THE VSTM FOR MULTI-FEATURE OBJECTS, WHICH WERE BASED ON THE DISCRETE-CAPACITY VIEWS, AND THE NEW QUESTIONS (Q1-Q3) RECONCEPTUALIZED BY USING THE CONTINUOUS-RESOURCE VIEW.....	161

Chapter 1
General Introduction

In a dynamic environment, it is important for an organism to store sensory information for later use. For example, a change in the environment might be caused by another organism, either a predator or a prey. Thus, possessing a mental representation of the recent past is important for an animal's survival and fitness (1). Memory takes many forms and spans many time scales, but we focus on the short-term buffer called short-term memory or working memory, in which sensory information is stored for potential use a few seconds later. Visual short-term memory (VSTM) is that buffer for visual information (2). In humans, visual short-term memory is not only useful to detect changes between two scenes (3), but also for integrating visual information between saccades (4–6). Humans make two to four saccadic eye movements every second (7), and these “snapshots” of a visual scene have to be integrated over time to provide a stable and unified view of our surroundings (8, 9). Finally, observers combine their visual memory with current visual information to guide hand movements (10).

1 History of research on VSTM

Behavioral studies of VSTM have a long history. Historically, investigation of the limitations of VSTM has been closely related to attempts to estimate the capacity of VSTM. In the late 19th century, Cattell (11) showed a series of letters briefly to observers, and found that they were able to recall up to about five letters. Kaufman et al. (12) found similar limitations using a “subitizing” paradigm, in which observers made a rapid, accurate and confident judgment of set size. The authors found that the speed, accuracy, and confidence of observers' judgments decreased when set size exceeded four. Several studies found that observers have limitations on tracking multiple objects simultaneously. Pylyshyn and

Storm (13) showed that observers could track and identify up to four items. This result was replicated in later studies (14, 15). Rensink (3) estimate capacity as 5..

In a different tradition, electrophysiologists have investigated the storage of information in VSTM in terms of neural activity. Fuster and Alexander (16) found that neurons in the monkey's prefrontal cortex and nucleus medialis dorsalis of the thalamus showed persistent firing during a delay period, which slowly decayed over time. Similar results were found in other studies (17–22). Kubota and Niki (17) found a group of neurons in the monkey prefrontal cortex that are exclusively active during the delay period. Funahashi et al. (18) reported that more than half of the prefrontal neurons (in the principal sulcus and frontal eye field) they recorded showed increased activity during the delay period. The activity of approximately 80% of these neurons was location-specific, which suggests that these neurons may hold information about stimuli during the delay period. Using face stimuli, Courtney (19) measured fMRI signals from human subjects and found that prefrontal areas showed sustained activity during a memory delay, in contrast to occipitotemporal areas, which showed transient activity after the stimuli.

Fuster (23) later suggested that sustained neural activity in VSTM may originate from a recurrent neural circuit (24), which may facilitate integration of information over time. This reverberation of neural activity has been suggested (5) to be related to transsaccadic integration (4).

2 History of main paradigms

Visual short-term memory has been examined using a variety of behavioral tasks. Here, I review three types of tasks: change detection, delayed estimation, and change localization.

2.1 Change detection

Change detection is an experimental paradigm that has been widely used in psychology for several decades. A typical change detection experiment consists of two sequential displays interleaved by a delay period also called inter-stimuli interval (ISI). At the end of each trial, observers make a decision on change occurrence of the trial. In a change detection task, the change can consist of the addition or deletion of an object from a scene (25, 26), or of a change in a feature of an item (27–37).

In earlier studies, change detection was used to distinguish VSTM from other visual memory systems. Phillips (30) used a change detection paradigm with unfamiliar visual stimuli. The author presented a black-and-white N by N matrix ($N = 4, 6, 8$) to subjects and varied the ISI. On half of the trials, one of the squares in the matrix changed from white to black or vice versa, and subjects decided whether a change occurred or not. With a short ISI (20 ms), subject performance was near perfect. However, the author found a significant effect of ISI (100-9000 ms) on performance: at longer ISIs, performance was lower and reaction times were longer. Moreover, subject performance gradually decreased when ISI further increased. The author later found that at 20 ms ISI, subject performance dramatically decreased when the first stimulus array was followed by a mask, although at a longer delay (≥ 100 ms), the memory was resistant to the presentation of a mask. The author identified this masking-resistant but capacity-limited memory as VSTM.

Pashler (38) used a change detection task with alphanumeric displays and found similar results as Phillips (30). The authors showed 10-letter displays briefly at varying ISIs. On half of the trials, the two displays were identical. On the other half, one of the

letters changed to a different letter. Subjects reported whether a change had occurred. The author found that with a shorter ISI (34 ms), performance was higher than longer ISIs. The performance gradually deteriorated as the ISI increased.

Luck and Vogel (28) found that performance in a color change detection task decreased with set size. Then, they introduced objects with multiple features: orientation, size, color, or spatial distance. They found that performance was independent of the number of features an object had.

2.2 Delayed estimation

Wilken and Ma (39) introduced a new paradigm called delayed estimation. In a color delayed estimation experiment, subjects viewed a brief display (100 ms) of N colored items ($N = 2, 4, 6$ or 8). After a delay (1500 ms), a cue was presented at the location of one of the previously presented items. The task was to report the color of the cued stimulus by clicking on one of the colors in a color wheel. The authors conducted similar experiments using orientation and spatial frequency. They found that the variance of the estimation error increased with set size in a monotonic fashion.

Delayed estimation differs from change detection in that experimenters may consider an observer's response as a proxy of their internal memory of stimulus.

2.3 Change localization

Change localization is similar to change detection except that exactly one change occurs on every trial and the task is to find the location of the change. A trial schematic example is shown in Fig. 1.1.

Buschman et al. (43) used this paradigm to investigate neural substrates of VSTM limitation in monkeys. They conducted a color change localization experiment with N items ($N = 2, 3, 4, \text{ or } 5$) and found that behavioral performance decreased with set size. Change localization performance is lower in people diagnosed with schizophrenia than in healthy adults (44).

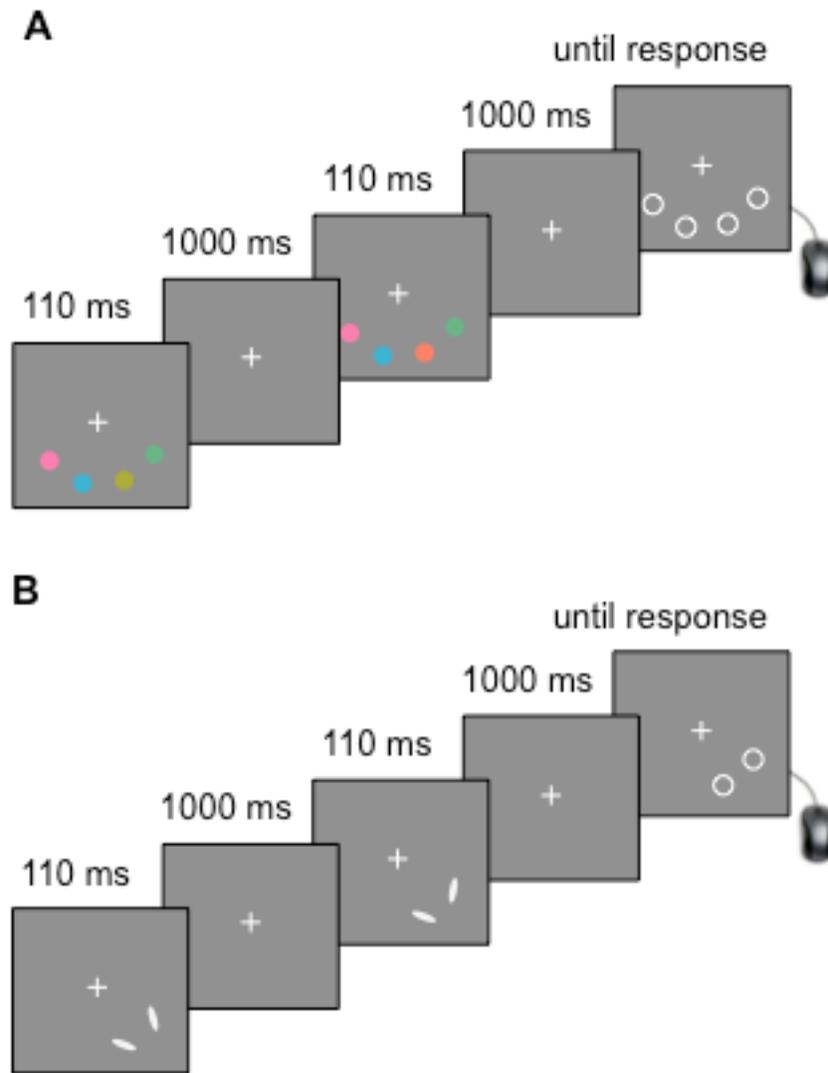


Figure 1.1 Trial procedure of change localization task. (A) Color change localization. (B) Orientation change localization.

3 History of VSTM models

It is widely agreed that people's VSTM performance decreases with set size (28, 38). However, qualitatively different models have been proposed for such set size effects.

3.1 Discrete-capacity models

The classical view on limitations of VSTM is that only a fixed number of items can be memorized. Miller's magic number seven (45) and Cowan's four (46) are examples. Several behavioral studies have reported a capacity of four in various VSTM tasks (11–13, 25, 28). Luck and Vogel (28) suggested that observers can memorize three or four items in object "slots," and if items are stored in the slots, observers have perfect memory of those items. The authors tested VSTM of orientation, color, spatial distance between stimuli, and size, as well as combinations of all features ("conjunctions"). Regardless of the type of features and the number of features in an object, the authors found that performance was only affected by the number of objects and performance dramatically decreased when set size was larger than four. This even applied to objects with two different colors ("bi-color" objects). The authors concluded that the unit of VSTM is an object, and that VSTM is limited by a fixed capacity of about four items.

There framework, the so-called "item-limit" view (the IL model), has garnered widespread following (28, 31, 47–50). However, the all-or-none encoding of items is difficult to motivate from a biological viewpoint.

To incorporate the notion of sensory noise, recently a modified version of the IL model, called the slots-plus-averaging (SA) model, was suggested (42). According to the model, the items "in memory" are encoded with a certain degree of precision. However, if

an item is not in memory, observers make a guess that follows a uniform distribution. The SA model allows for noisy encoding of stimuli, but still assumes that VSTM consists of discrete memory packages, so-called “slots”, which can be either stacked in or distributed across certain items. The number of slots determines how many times a stimulus is sampled. If a stimulus receives multiple slots, it is sampled multiple times, which increases precision. Similar to the IL model, when set size is larger than the number of slots, some items cannot be encoded and observers have to make a guess about them.

3.2 Continuous-resource models

The discrete-capacity view has been recently questioned by researchers who suggested that VSTM is best represented as a shared pool of a continuous resource. The origin of this idea is the sample size model (29, 53), in which visual attention is represented by a large but fixed number of samples equally distributed over all items. Thus, each item receives a number of samples that is inversely proportional to set size. Each item is represented as an internal measurement with noise, which is determined by the number of samples it receives.

Wilken and Ma (39) used a delayed estimation paradigm and showed memory precision decreased systematically with set size. Bays and Husain (40) suggested that the limited resource of VSTM is highly flexible, and thus the resource needs to be allocated to a specific location based on its attentional demand, which can be further updated. Both studies propose that every item is encoded with a certain degree of precision and supports the resource concept of VSTM. Recent human psychophysics have supported the resource concept of VSTM (35, 36, 54, 55). An attentional tracking task with a model comparison found that discrete-capacity models failed to describe subject performance (55).

In many aspects, the continuous-resource models have a more solid biological foundation than the discrete-capacity models. The idea of flexible allocation of resource can be related to physiological findings such as trial-to-trial variability in attentional gain (56, 57). Finally, a recent primate study showed that encoding precision in VSTM decreases with set size in change detection (58), and another primate study revealed that prefrontal cortex predominantly encodes attentional allocation (59). Thus, the continuous-resource models pose a serious challenge to the discrete-capacity models.

Figure 1.1 summarizes three previously suggested and a new VSTM model, which will be introduced in the next chapter. The IL and SA models are both discrete-capacity models. The equal-precision (EP) model is similar to Shaw's sample size model (53) except that VSTM is a continuous form of resource instead of discretized samples.

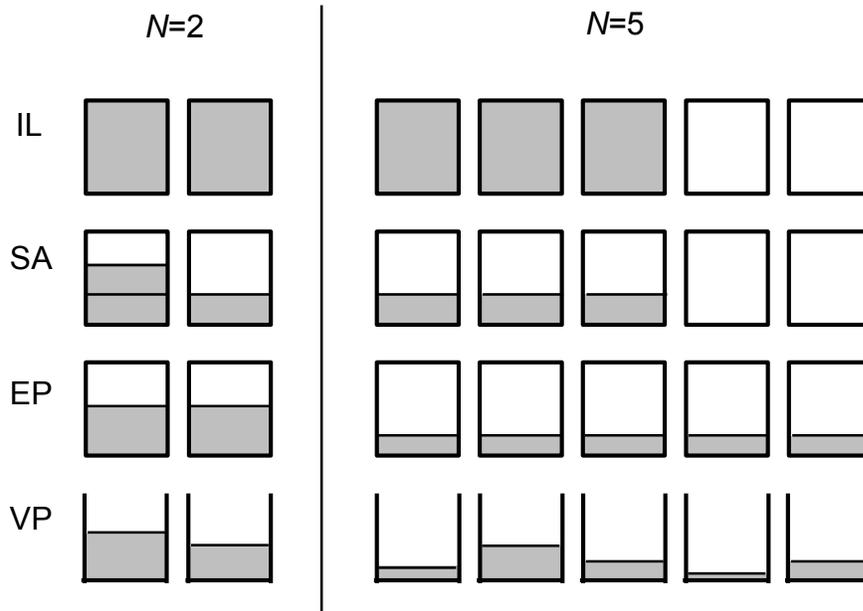


Figure 1.2 Models of VSTM (IL: the item-limit model, SA: the slots-plus-averaging model, EP: the equal-precision model, VP: the variable-precision model). Each box represents an item. Capacity or resource is represented in gray. Set size is 2 (left) or 5 (right). In this example, the number of “slots” or “chunks” is 3 in the IL and SA models.

4 Overview of chapters

In this thesis, I attempt to establish a new framework of VSTM limitations. I propose a new model and compare the predictions of this model with previously suggested models. In the rest of the thesis, I revisit some aspects of VSTM limitations within the context of the new framework.

In Chapter 2, I propose an alternative model of VSTM limitations, namely the variable-precision (VP) model, and compare the predictions of the model with other models that were suggested in previous literature. In Chapter 3, I explore a richer space of models. In Chapters 4 and 5, I study VSTM for multi-feature objects within the continuous-resource view. In Chapter 6, I examine VSTM as a function of delay time.

Chapter 2

Modeling set-size effects in VSTM

(Portions of) this work have been published in R van den Berg*, H Shin*, WC Chou, R George, WJ Ma (*:joint first authors). (2012) Variability in encoding precision accounts for visual short-term memory limitations, *PNAS* 109 (22), 8780-878.

1 Introduction

Cognitive psychology has traditionally conceptualized visual short-term memory (VSTM) as limited by a fixed number of items, called the capacity. In this view, each item is encoded in an all-or-none fashion; when the number of items given to an observer to memorize, called set size, exceeds the capacity, some items will not be stored. Much more recently, it has been hypothesized that VSTM might store all items under all circumstances, but with lower precision per item as set size increases. After reviewing previous models, I here propose a specific instantiation of this idea – the variable-precision (VP) model – that makes the entire concept of precision-based models empirically viable. To compare the VP model against the classic model and other alternative models, I conducted two human psychophysics experiments (Experiment 1: color, and Experiment 2: orientation) in which observers localize where in a scene a change occurred, and the number of items in the scene changes from trial to trial. I found that the VP model provides a better account of the behavior of every subject than the other models, in both experiments.

2 Summary of task

Observers are shown two visual displays separated by 1 s delay (Fig. 1.1). Between the displays, a change occurs in one feature of one item. Observers click on the location where the change occurred. The first display contains N items that vary along one feature dimension (either orientation or color). Feature values are drawn independently across items from a uniform distribution over a circle in that feature space (all possible orientations for orientation; color wheel for color). The feature value of the changed

object is also independently drawn from the same uniform distribution. Each location has the same probability of containing the change.

3 Theory and models

I model the observer's change localization decisions using a Bayes-optimal model. A Bayesian model consists of three steps: 1) the generative model (encoding model), which describes the statistics of the variables in the task, 2) the inference (decision-making) model, which describes how an observer reaches a decision based on their observations on a given trial, and 3) a calculation of response probabilities, i.e., how often the model predicts the observer to make each possible response on a given trial.

3.1 Step 1: Encoding: Measurement

3.1.1 Noise

A physical stimulus elicits activity in the nervous system. This activity will vary randomly from trial to trial even when the physical stimulus itself is identical each time. Such variability originates from many sources. Our sensors are subject to random variability due to intrinsic stochastic processes. For instance, thermal noise affects the responses of hair cells in the inner ear that sense sound waves. The transduction process by which the nervous system captures physical energy and converts it into an electrical response is also stochastic. For instance, the absorption of photons by photoreceptors is a stochastic process; only sometimes is there a response to a single photon. At the subcellular level, neurotransmitter release and ion channel opening and closing are stochastic processes.

In some cases, this noise can be easily illustrated. For example, if we place the index finger of our right hand on top of a table and try to place the index finger of our left hand at the matching location underneath the table, we often observe quite a difference (typical variability is about 2 cm in this task). This indicates noise in our internal proprioceptive representations of limb location. Similarly, it is difficult to estimate whether one object is heavier than another based on our sense of force because the internal distribution of force is noisy – necessitating the use of scales to compare weights. Measurement of the depth of an object is also subject to considerable noise, particularly with monocular viewing. One can easily test this by trying to bring the tips of his or her index fingers together while one eye is closed. These examples illustrate that the relationship between stimulus and sensor response is stochastic.

A *measurement* can be defined as the representation in stimulus space of the internal neural representation of the stimulus. For example, if the true location s of the sound is 3° to the right of straight ahead, then its measurement x could be 2.7° or 3.1° . The terminology “measurement” stems from the analogy with making physical measurements. If a stick is 89.0 cm long, you might measure its length to be 89.5, 88.1, 88.9 cm, or so on. The measurement lives in the same space as the stimulus, because it has the same units as the stimulus.

3.1.2 *Gaussian assumption*

The measurement distribution characterizes the measurement x for a given stimulus value s . This conditional distribution, $p(x|s)$, describes the frequency of occurrence of each value of the measurement when the same stimulus value s is repeated many times (60).

If many sources contribute to the variability of the measurement, a measurement distribution becomes roughly Gaussian. This assertion is – loosely – a consequence of the Central Limit Theorem. The Gaussian form of the measurement distribution is quite fundamental, independent of the experimental design. Thus, the equation for the measurement distribution is

$$p(x|s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-s)^2}{2\sigma^2}} \quad (2.1)$$

where σ is the standard deviation of the noise in the measurement, also called measurement noise level or sensory noise level. The higher σ , the noisier the measurement and the wider its distribution. Lowering the light level in a room, increasing the distance to an object, or removing your corrective eyewear are all ways to increase the standard deviation of the measurement noise of a visual stimulus. For an auditory or tactile stimulus, the same is achieved by introducing background noise, decreasing the intensity of a stimulus, or presenting it for a shorter time. The inverse of the variance of the measurement distribution, $1/\sigma^2$, is sometimes called the *reliability* or the *precision* of the measurement x .

3.1.3 Von Mises distribution

A slight complication arises from the fact that the stimulus spaces I use (orientation and color) are circular, so that the Gaussian distribution is no longer appropriate. For instance, if an orientation stimulus of 355° is encoded with 10° of noise, an observer's measurement can be 365° , which is the same as 5° . Therefore, I assume that the measurement follows a Von Mises distribution instead:

$$p(x|s) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-s)} \equiv \text{VM}(x; s, \kappa).$$

I_0 is the modified Bessel function of the first kind of order zero (61) and serves as a normalization. The concentration parameter κ controls the width of the noise distribution. When it is large, the Von Mises distribution resembles a Gaussian distribution with variance $1/\kappa$. When $\kappa=0$, $p(x|s)$ is the uniform distribution.

3.2 Step 1: Encoding: Precision and resource

Intuitively, resource is something that is allocated to an item to improve the quality of its encoding. The traditional notion of resource is that of a very large pool of available observations made of the stimulus, also called samples (29). Each observation is corrupted by independent, zero-mean Gaussian noise with the same standard deviation, and the observer's eventual measurement, x , is the mean of these observations. Then the variance of the measurement decreases in inverse proportion to the number of observations, and precision increases proportionally.

Here, I instead identify resource with *Fisher information*, denoted J . Fisher information determines the best possible performance of any estimator, through the Cramér-Rao bound (62). Fisher information is defined in terms of the noise distribution, which is the distribution of the observations conditioned on the stimulus s :

$$J(s) = - \left\langle \frac{\partial^2}{\partial s^2} \log p(\text{observations} | s) \right\rangle, \quad (2.2)$$

where $\langle \rangle$ denotes an expected value over $p(\text{observations}|s)$.

If x follows a Gaussian distribution with mean s and standard deviation σ , it is easily verified from the definition, Eq. (2.2), that Fisher information is equal to the inverse variance, $J = 1/\sigma^2$, covering the earlier relationship. This equation is an improvement over the “number of observations” argument because J is defined on a continuum and readily neurally interpretable. At the neural level, Fisher information is proportional to the gain of a population when neural variability is Poisson-like (63).

I calculate Fisher information from its definition, Eq. (1.2):

$$J = \langle \kappa \cos(x-s) \rangle = \frac{\kappa}{2\pi I_0(\kappa)} \int \cos(x-s) e^{\kappa \cos(x-s)} dx = \kappa \frac{I_1(\kappa)}{I_0(\kappa)}, \quad (2.3)$$

where $I_1(\kappa)$ is the modified Bessel function of the first kind of order one (61). This equation relates Fisher information in a one-to-one fashion to the concentration parameter of the Von Mises distribution. I use it in all models except for the IL model. One can think of Fisher information as precision, by analogy to the Gaussian case. I write the inverse relationship of Eq. (2.3) as

$$\kappa = \Phi(J). \quad (2.4)$$

The inverse function Φ is not analytical but can be computed numerically.

3.3 Step 1: Encoding: Models of memory

In the item-limit (IL), model, an item is encoded either perfectly or not at all. All other models I tested contain a notion of noise. Therefore, I have to specify the relationship between “amount of resource” and the level of noise.

3.3.1 Equal-precision model

In the equal-precision (EP) model, I assume

$$J = \frac{J_1}{N^\alpha}, \quad (2.5)$$

where J_1 is the Fisher information at set size 1, and α is a power parameter, which determines the power of set size effect. If α is 0, precision is not affected by set size. Using Eq.(2.4), the concentration parameter at set size N is

$$\kappa(N) = \Phi\left(\frac{J_1}{N^\alpha}\right). \quad (2.6)$$

3.3.2 Slots-and-averaging model

The slots-plus-averaging (SA) model (42) is similar to the IL model, with the modification that when $N < K$, multiple chunks of resource can be assigned to a single item. This modification gives it some characteristics of the EP model. Specifically, the assumption is that the amount of resource is proportional to the number of assigned chunks, S . Zhang and Luck (42) do not mention the exact relationship between amount of resource and the concentration parameter of the Von Mises distribution, but I will assume that they used the correct relationship, Eq. (2.3). Then, the concentration parameter as a function of S is

$$\kappa = \Phi(SJ_1), \quad (2.7)$$

where J_1 is now the Fisher information corresponding to having one slot ($S=1$). When $N > K$, an item receives 0 chunks or 1 chunk of memory, with probabilities K/N and $1-K/N$, respectively. This is the same as in the IL model. When $N \leq K$, all items receive at least one chunk and it is assumed that the chunks are distributed as equally as possible over all items.

For example, if $K=4$ and $N=3$, two items get assigned one chunk each and one item gets two chunks. From this, it follows that the number of chunks an item receives, S , is equal to

$$S = \begin{cases} \left\lfloor \frac{K}{N} \right\rfloor & \text{with probability } 1 - \frac{K \bmod N}{N}; \\ \left\lfloor \frac{K}{N} \right\rfloor + 1 & \text{with probability } \frac{K \bmod N}{N}, \end{cases}$$

where $\lfloor x \rfloor$ denotes the largest integer smaller than x (floor function). Using Eq. (2.7), these two values of S correspond to two values of the concentration parameter κ , which I denote by κ_{low} and κ_{high} , respectively:

$$\begin{aligned} \kappa_{\text{low}}(N) &= \Phi\left(\left\lfloor \frac{K}{N} \right\rfloor J_1\right), \\ \kappa_{\text{high}}(N) &= \Phi\left(\left(\left\lfloor \frac{K}{N} \right\rfloor + 1\right) J_1\right). \end{aligned} \tag{2.8}$$

In the example above, two items would be memorized with concentration parameter κ_{low} and the third one with κ_{high} .

3.3.3 Variable-precision model

In the variable-precision model, precision is variable across items and trials. I assume that each precision is drawn independently from a gamma distribution with mean precision \bar{J} and scale parameter τ (Fig. 2.1A),

$$p(J | \bar{J}; \tau) = \text{Gamma}(J; \bar{J}, \tau), \tag{2.9}$$

The measurement is then described by a doubly stochastic process, $(\bar{J}, \tau) \rightarrow J \rightarrow x$. The variance of J is equal to $\bar{J}\tau$. The gamma distribution is a common distribution on the

positive real line. I assume that mean precision, \bar{J} , depends on set size in the following way (Fig. 2.1B):

$$\bar{J} = \frac{\bar{J}_1}{N^\alpha}. \quad (2.10)$$

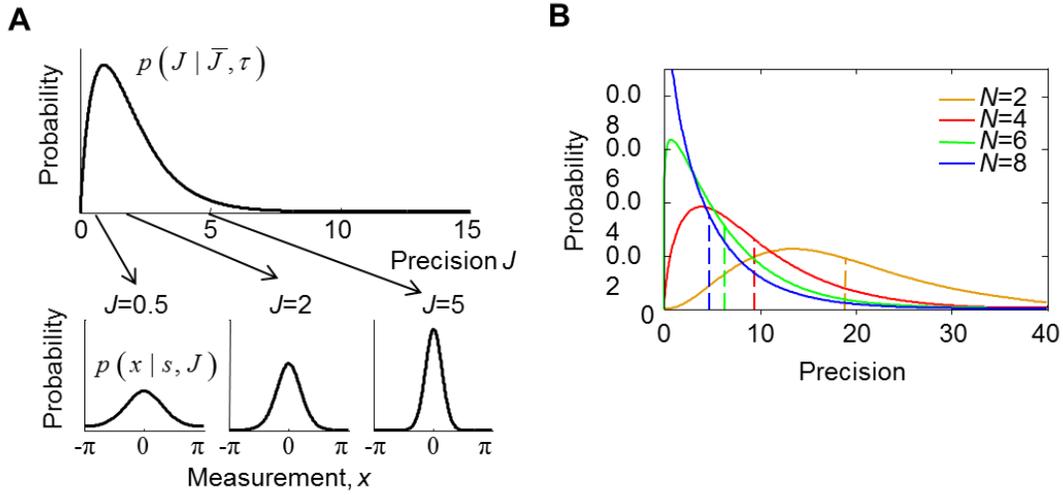


Figure 2.1 Theory. (A) Top: In the VP model, precision, J , is variable and assumed to follow a gamma distribution (here with $\tau=1$). Bottom: Von Mises noise distributions corresponding to three values of precision and $s=0$. A higher J produces a narrower distribution $p(x|s,J)$. (B) Example probability distributions over precision at different set sizes in the VP model. Here, mean precision (dashed lines) was taken inversely proportional to set size ($\alpha=1$). In the EP model, these distributions would be delta functions.

3.4 Step 2: Inference

In change localization, the variables in the task are the location of the change, L , the magnitude of the change, Δ , the vector of stimuli in the first display, $\boldsymbol{\theta}=(\theta_1,\dots, \theta_N)$, and the vector of stimuli in the second display, $\boldsymbol{\varphi}=(\varphi_1,\dots, \varphi_N)$. Each L has a probability of $1/N$. The probability density of Δ is flat at $\frac{1}{2\pi}$ ($\frac{1}{\pi}$ for orientation here and the following equations) and the one over $\boldsymbol{\theta}$ flat at $\left(\frac{1}{2\pi}\right)^N$. The relation between $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ is $\boldsymbol{\varphi}=\boldsymbol{\theta}+\Delta\mathbf{1}_L$, where $\mathbf{1}_L$ is the vector of zeroes with a 1 at the L^{th} entry.

All models except for the IL model have noise in the measurements and probabilistic inference is needed to estimate the location of the change. The i^{th} measurement in the first display, x_i , is drawn independently from a Von Mises distribution with mean θ_i and concentration parameter κ_i . The i^{th} measurement in the second display, denoted y_i , is drawn from a Von Mises distribution with mean φ_i and concentration parameter κ_i (it is possible to allow κ_i to be different between the two displays but I chose not to do so). The relations between the variables are shown in the graphical model in Fig. 4. To model how the observers decides based on the measurements $\mathbf{x}=(x_1,\dots,x_N)$ and $\mathbf{y}=(y_1,\dots,y_N)$, I use a Bayesian-observer model. The Bayesian observer computes a probability distribution over the location of the change, $p(L|\mathbf{x},\mathbf{y})$, and then reports the location with the highest probability. The posterior distribution over L is proportional to the joint distribution, $p(\mathbf{x},\mathbf{y},L)$, which in turn is evaluated as an integral over the remaining variables, namely Δ , $\boldsymbol{\theta}$, and $\boldsymbol{\varphi}$:

$$\begin{aligned}
p(\mathbf{x}, \mathbf{y}, L) &= \iiint p(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \Delta, L) d\Delta d\boldsymbol{\theta} d\boldsymbol{\varphi} \\
&= \iiint p(L) p(\Delta) p(\boldsymbol{\theta}) p(\boldsymbol{\varphi} | L, \boldsymbol{\theta}) p(\mathbf{x} | \boldsymbol{\theta}) p(\mathbf{y} | \boldsymbol{\varphi}) d\Delta d\boldsymbol{\theta} d\boldsymbol{\varphi},
\end{aligned}$$

where in going from the first to the second line I have used the structure of the generative model in Fig. 2.2. Substituting distributions and evaluating the integral over $\boldsymbol{\varphi}$ gives

$$p(\mathbf{x}, \mathbf{y}, L) = \frac{1}{N} \left(\frac{1}{2\pi} \right)^{N+1} \int \prod_{i=1}^N \left(\int p(x_i | \theta_i) p(y_i | \varphi_i = \theta_i + \Delta \delta_{L,i}) \right) d\Delta, \quad (2.11)$$

where $\delta_{L,i}=1$ when $L=i$ and 0 otherwise. Since I am only interested in the dependence on L ,

I can freely divide by the L -independent product $\prod_{i=1}^N \left(\int p(x_i | \theta_i) p(y_i | \varphi_i = \theta_i) \right)$, leaving

only integrals pertaining to the L^{th} location:

$$p(\mathbf{x}, \mathbf{y}, L) \propto \frac{\iint p(x_L | \theta_L) p(y_L | \varphi_L = \theta_L + \Delta) d\theta_L d\Delta}{\int p(x_L | \theta_L) p(y_L | \varphi_L = \theta_L)}. \quad (2.12)$$

This evaluates to

$$p(\mathbf{x}, \mathbf{y}, L) \propto \frac{I_0(\kappa_L)^2}{I_0\left(\kappa_L \sqrt{2 + 2 \cos(x_L - y_L)}\right)}.$$

Thus, the maximum-a-posteriori (MAP) estimate of the location of the change is

$$\hat{L} = \underset{L}{\operatorname{argmax}} \frac{I_0(\kappa_L)^2}{I_0\left(\kappa_L \sqrt{2 + 2 \cos(x_L - y_L)}\right)}. \quad (2.13)$$

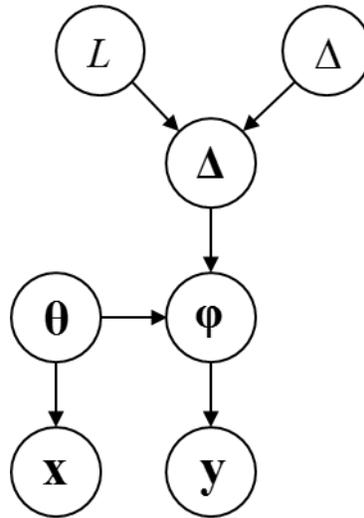


Figure 2.2 Generative model for the change localization task. L : location of a change; Δ : magnitude of change; Δ : vector of change magnitudes at all locations; θ and φ : vectors of stimuli in the first and second displays, respectively; x and y : vectors of measurements in the first and second displays, respectively.

3.5 Step 3: Response probabilities

In all models, each trial is characterized completely by its change magnitude Δ ; the values of stimuli θ and ϕ do not matter otherwise. Moreover, the observer's response is completely characterized by whether it is correct or not; all incorrect responses are equivalent. For a given parameter combination, ω , and a given change magnitude Δ , I calculated the probability of a correct response, denoted by $p(\text{correct}|\Delta, \omega)$, through Monte Carlo simulation. This entailed the following. I generated 10,000 samples of $J_{x,i}$ and $J_{y,i}$ for the first and second stimulus array according to the process described under Step 1. I used these values to compute concentration parameters $\kappa_{x,i}$ and $\kappa_{y,i}$ according to Eq. (2.3), and I drew measurement vectors \mathbf{x} and \mathbf{y} from Von Mises distributions with those concentration parameters, respectively. Those von Mises distributions all had mean 0 except that one element of \mathbf{y} was drawn from a Von Mises distribution with mean Δ . For each of the 10,000 combinations of κ_x , κ_y , \mathbf{x} , and \mathbf{y} thus drawn, I evaluated the decision rule, Eq. (2.13). Tallying correctness across these draws yielded an estimate of the probability of a correct localization response for the given parameter combination and the given change magnitude. I used these probabilities for model fitting.

3.5.1 Equal-precision model

In the equal-precision model, I take $\kappa_i = \Phi\left(\frac{J_i}{N^\alpha}\right)$ for all i . I assume equality between first and second displays, because the concentration parameters in both displays can essentially not be fitted independently (compare to the sum of two normally distributed random

variables: the variances sum and cannot be estimated individually). The EP model has two free parameters: J_1 and α .

3.5.2 Slots-and-averaging model

In the slots-plus-averaging model, κ_i is given by Eq. (2.8). When $N \leq K$, Eq. (2.13) applies. When $N > K$, inference is based on only K out of N measurements in each display, $\mathbf{x}=(x_1, \dots, x_K)$ and $\mathbf{y}=(y_1, \dots, y_K)$, yet the change could have occurred at any location. I first evaluate the joint probability of \mathbf{x} , \mathbf{y} , and that the change occurred at a location L that is among the encoded ones. In analogy to Eq. (2.11), this probability is

$$(L \text{ encoded}) \quad p(\mathbf{x}, \mathbf{y}, L) = \frac{1}{N} \left(\frac{1}{2\pi} \right)^{K+1} \int \prod_{i=1}^K \left(\int p(x_i | \theta_i) p(y_i | \varphi_i = \theta_i + \Delta \delta_{L,i}) \right) d\Delta. \quad (2.14)$$

Now I evaluate the joint probability of \mathbf{x} , \mathbf{y} , and that the change occurred at a location L that is not among the encoded ones. This is equal to

$$\begin{aligned} (L \text{ not encoded}) \quad p(\mathbf{x}, \mathbf{y}, L) &= \iint p(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\varphi}, L) d\boldsymbol{\theta} d\boldsymbol{\varphi} \\ &= \iint p(L) p(\boldsymbol{\theta}) p(\boldsymbol{\varphi} | L, \boldsymbol{\theta}) p(\mathbf{x} | \boldsymbol{\theta}) p(\mathbf{y} | \boldsymbol{\varphi}) d\boldsymbol{\theta} d\boldsymbol{\varphi} \quad (2.15) \\ &= \frac{1}{N} \left(\frac{1}{2\pi} \right)^K \prod_{i=1}^K \left(\int p(x_i | \theta_i) p(y_i | \varphi_i = \theta_i) \right). \end{aligned}$$

As one would expect, this probability does not depend on L . Since I am only interested in the location L for which $p(\mathbf{x}, \mathbf{y}, L)$ is largest (i.e. the argmax), I divide both Eqs. (2.14) and (2.15) by Eq. (2.15). Then, in analogy to Eq. (2.12), I have to take the argmax of

$$\begin{cases} (L \text{ encoded}) & \frac{1}{2\pi} \frac{\iint p(x_L | \theta_L) p(y_L | \varphi_L = \theta_L + \Delta) d\theta_L d\Delta}{\int p(x_L | \theta_L) p(y_L | \varphi_L = \theta_L)} = \frac{1}{2\pi \int p(x_L | \theta_L) p(y_L | \varphi_L = \theta_L)} \\ (L \text{ not encoded}) & 1 \end{cases}$$

Evaluating the integral, the estimate of location is

$$\hat{L} = \underset{L}{\operatorname{argmax}} \frac{I_0(\kappa_L)^2}{I_0\left(\kappa_L \sqrt{2 + 2 \cos(x_L - y_L)}\right)} \quad (2.16)$$

when the value of this maximum exceeds 1, and a random guess from among the non-encoded items when it does not. The SA model has two free parameters: J_1 and K .

3.5.3 Variable-precision model

In the variable-precision model, every J_i is independently drawn from a gamma distribution with mean \bar{J} (given by Eq. (2.10)) and scale parameter τ . Then, the estimate distribution is

$$p(\hat{L} | L, \Delta; \bar{J}, \tau) = \int \cdots \int p(\hat{L} | L, \Delta; \mathbf{J}) \left(\prod_{i=1}^N \text{Gamma}(J_i; \bar{J}, \tau) \right) dJ_1 \cdots dJ_N.$$

This distribution is obtained through Monte Carlo simulation of \mathbf{J} , using 10,000 samples.

The VP model has three free parameters: \bar{J}_1 , α , and τ .

3.6 Item-limit model

According to the item-limit model, the probability of being correct is equal to $1 - \varepsilon$ when

$N \leq K$ and to $\frac{K}{N} + \left(1 - \frac{K}{N}\right) \frac{1}{N - K} = \frac{K + 1}{N}$ when $N > K$. This is independent of θ , ϕ , and Δ . I

introduced ε because without it (i.e. $\varepsilon = 0$), the data would have probability zero under the model. The IL model has two free parameters: K and ε .

4 Methods

4.1 Psychophysics

4.1.1 *Experiment 1: Change localization with color stimuli*

Observers briefly viewed two screens containing a set of colors, separated in time by a blank screen, and reported the location of the color change (Fig. 1.1A).

4.1.1.1 Stimuli

All stimuli were displayed on a 19" LCD monitor at a viewing distance of approximately 60 cm. The first stimulus array was composed of N colored discs ($N=2, 4, 6, \text{ or } 8$) with a diameter of 0.62 degrees of visual angle (deg) with their centers lying on an imaginary circle of radius 7 deg (Fig. 2C). The locations of the discs were randomly selected from eight fixed positions equally spaced along the circle, including the positions corresponding to the cardinal directions with respect to fixation. The colors were drawn independently from 180 color values uniformly distributed along a circle of radius 60 in CIE 1976 (L^* , a^* , b^*) color space. This circle had constant luminance ($L^*=58$) and was centered at the point ($a^*=12, b^*=13$). The stimuli were presented on a mid-level grey background (128 on an 8-bit greyscale) of luminosity 33.1 cd/m^2 .

4.1.1.2 Procedure

The trial sequence consisted of the presentation of a fixation cross (1000 ms), the stimulus array (110 ms), a delay period during which only the fixation cross was visible (1000 ms), another stimulus array in which one of the stimuli changed color (110 ms), and a response screen that consisted of empty circles at the locations where the stimuli were shown. In the

first stimulus array, set size was chosen randomly and the color of each item was chosen randomly as described above. In the second stimulus array, $N-1$ stimuli were identical to those in the first display, while the color of the remaining stimulus was chosen randomly from the same uniform distribution. The location of the changing stimulus was chosen randomly. The subject's task was to click on the location of the stimulus that had changed color. The experiment consisted of 4 sessions on different days. Each session consisted of 4 blocks with 120 trials each. Hence, each subject completed $4 \cdot 4 \cdot 120 = 1920$ trials in total. Seven subjects participated in this experiment (age range 21-32 years; 5 naïve).

4.1.2 Experiment 2: Change localization with orientation stimuli

Experiment 2 differed from Experiment 1 in the following ways. Stimuli were white, oriented ellipses with minor and major axes of 0.41 and 0.94 degrees of visual angle (deg), respectively, and a luminance of 95.7 cd/m^2 (Fig. 1.1B). Eleven subjects participated (age range 23-29 years; 9 naïve).

4.2 Model comparison

Bayesian model comparison is a powerful method to compare models, because it can use individual-trial responses instead of summary statistics, and because it automatically penalizes models with more free parameters. Each model m produces a predicted probability distribution of a correct response, denoted by $p(\text{correctness} | m; \Delta, N, \omega)$, where Δ is a magnitude of change, N is a set size, and ω is a parameter vector. Bayesian model comparison consists of calculating for each model the probability of finding a subject's actual responses under this distribution, averaged over free parameters:

$$\begin{aligned}
L(m) &= p(\text{data} | m) \\
&= \int p(\text{data} | m, \boldsymbol{\omega}) p(\boldsymbol{\omega} | m) d\boldsymbol{\omega} \\
&= \int \left(\prod_{t=1}^{N_{\text{trials}}} p(\text{correctness}_t | m; \Delta_t, N_t, \boldsymbol{\omega}) \right) p(\boldsymbol{\omega} | m) d\boldsymbol{\omega},
\end{aligned}$$

It is convenient to take the logarithm and rewrite as

$$\log L(m) = \log L_{\max}(m) + \log \int \exp(\log L(m; \boldsymbol{\omega}) - \log L_{\max}(m)) p(\boldsymbol{\omega} | m) d\boldsymbol{\omega}, \quad (2.17)$$

where $\log L(m; \boldsymbol{\omega}) = \sum_{t=1}^{N_{\text{trials}}} \log p(m; \Delta_t, N_t, \boldsymbol{\omega})$ and $L_{\max}(m) = \max_{\boldsymbol{\omega}} L(m; \boldsymbol{\omega})$. This form

prevents numerical problems, since the exponential in the integrand of Eq. (2.17) is now of order 1 near the maximum-likelihood value of $\boldsymbol{\omega}$. For the prior, I assume a uniform distribution across a plausible range (see Table 2.1), whose size I denote S_j for the j^{th} parameter. Then Eq. (2.17) becomes

$$\log L(m) = \log L_{\max}(m) - \sum_{j=1}^{\dim \boldsymbol{\omega}} \log S_j + \log \int \exp(\log L(m; \boldsymbol{\omega}) - \log L_{\max}(m)) d\boldsymbol{\omega},$$

where $\dim \boldsymbol{\omega}$ is the number of parameters. I approximated the integral through a Riemann sum, with 25 bins in each parameter dimension. The ratio of likelihoods of two models is also known as a Bayes factor. As an alternative to Bayesian model comparison, the

Bayesian information criterion is $\text{BIC} = \log L_{\max} - \frac{\dim \boldsymbol{\omega}}{2} \log N_{\text{trials}}$.

5 Results

5.1 Summary statistics

Set size had a significant main effect on accuracy both for color (one-way repeated-measures ANOVA, $F(3,18)=256.6$, $p<0.001$) and orientation ($F(3,30)=356.5$, $p<0.001$); see Fig. S8A and S9A. Magnitude of change has a significant effect on accuracy both for color (one-way repeated-measures ANOVA, $F(8,48)=114.3$, $p<0.001$) and orientation ($F(8,80)=238.5$, $p<0.001$); see Fig. 6. Judged by RMS error, the VP model provides the best fits to the psychometric curves (Fig. 2.3). Individual-subject fits are shown in Figs. S8-9. In the SA model, capacity K equals 2.86 ± 0.14 for color and 4.09 ± 0.39 for orientation. In the VP model, the power α equals 0.974 ± 0.090 for color and 0.993 ± 0.075 for orientation. Maximum-likelihood estimates of the parameters in all models in all experiments are shown in Table 2.1.

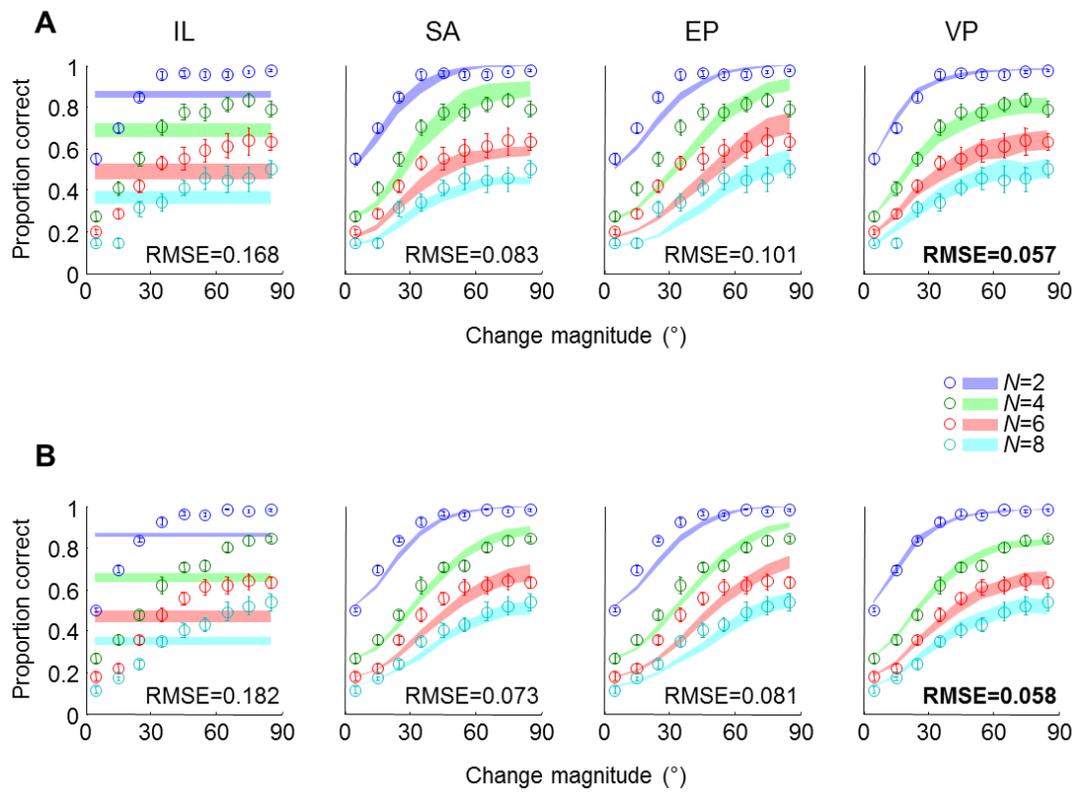


Figure 2.3 Proportion correct (subject average) as a function of change magnitude at each set size (N) in the color (A) and orientation experiments (B).

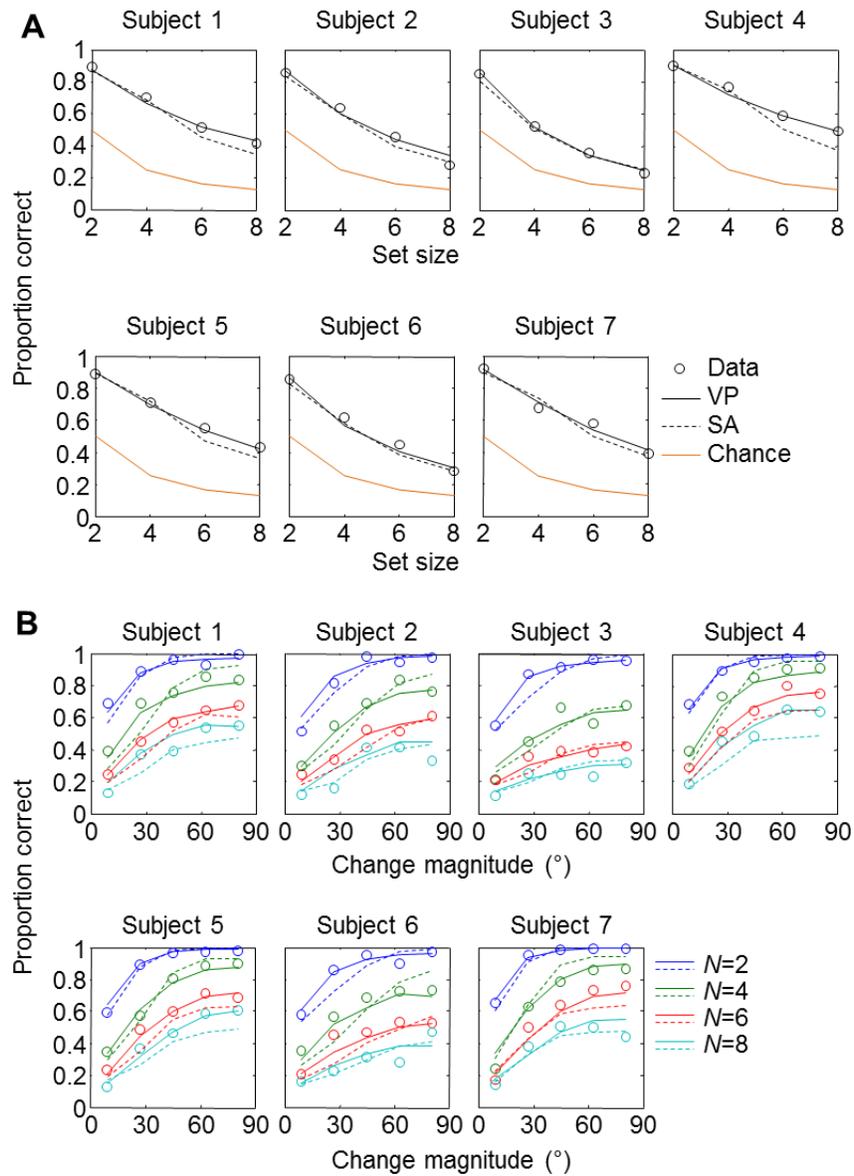


Figure 2.4 Individual-subject fits of the SA and VP models in the color experiment (solid: VP; dashed: SA; other models are not shown to avoid clutter). (A) Proportion correct as a function of set size. (B) Proportion correct as a function of change magnitude at each set size.

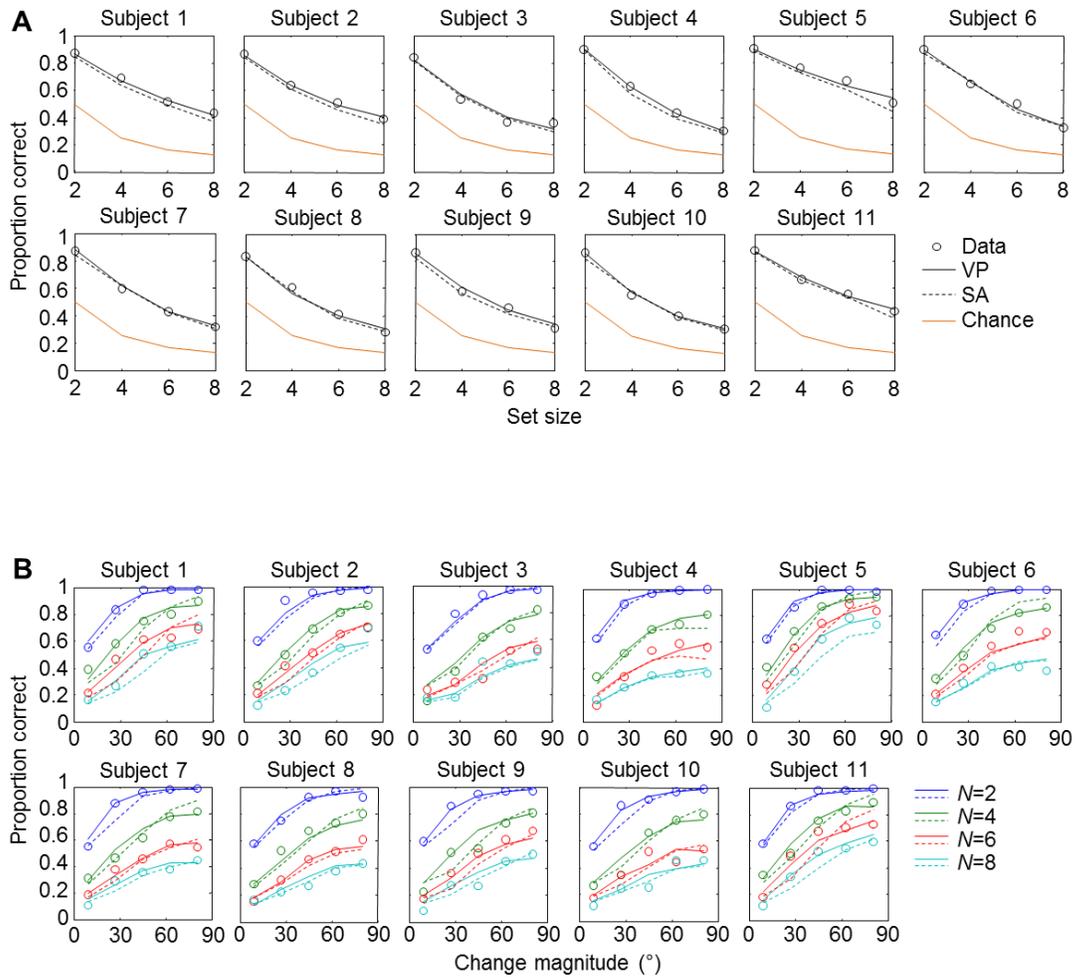


Figure 2.5 Individual-subject fits of the SA and VP models in the orientation experiment (solid: VP; dashed: SA; other models are not shown to avoid clutter). (A) Proportion correct as a function of set size. (B) Proportion correct as a function of change magnitude at each set size.

Model	Parameter	Experiment 1		Experiment 2		Tested range	
		Mean	S.e.m.	Mean	S.e.m.	Min	Max
IL	K	2.71	0.36	2.55	0.28	1	8
	ε	0.279	0.035	0.282	0.021	0	1
SA	K	2.86	0.14	4.09	0.39	1	8
	J_1	5.9	1.0	3.51	0.72	1	15
EP	J_1	11.2	1.8	11.1	1.6	1	25
	α	1.018	0.067	0.984	0.071	0	2
VP	\bar{J}_1	42.3	3.3	27.4	4.5	1	60
	α	0.974	0.090	0.993	0.075	0	2
	τ	36.0	5.7	13.1	1.9	1	60

Table 2.1 Means and SEs of the maximum-likelihood estimates and tested ranges of model parameters.

5.2 Model comparison

In Bayesian model comparison, the VP model outperforms the IL, SA, and EP both for color (by 143 ± 11 , 10.1 ± 2.6 , and 15.0 ± 2.8 log likelihood points) and for orientation (by 145 ± 11 , 11.9 ± 2.6 , and 17.3 ± 2.8 points) (Fig. 2.6). In both experiments, the VP model outperforms all other models for every individual subject (Figs. 2.7 and 2.8).

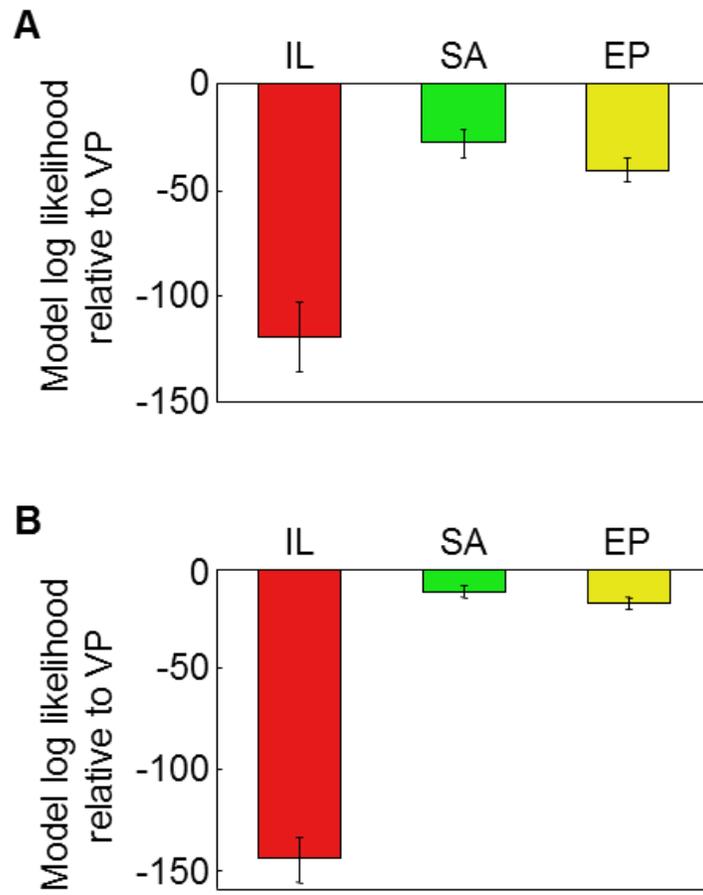


Figure 2.6 Model log likelihoods relative to the VP model. (A) Color experiment. (B) Orientation experiment.

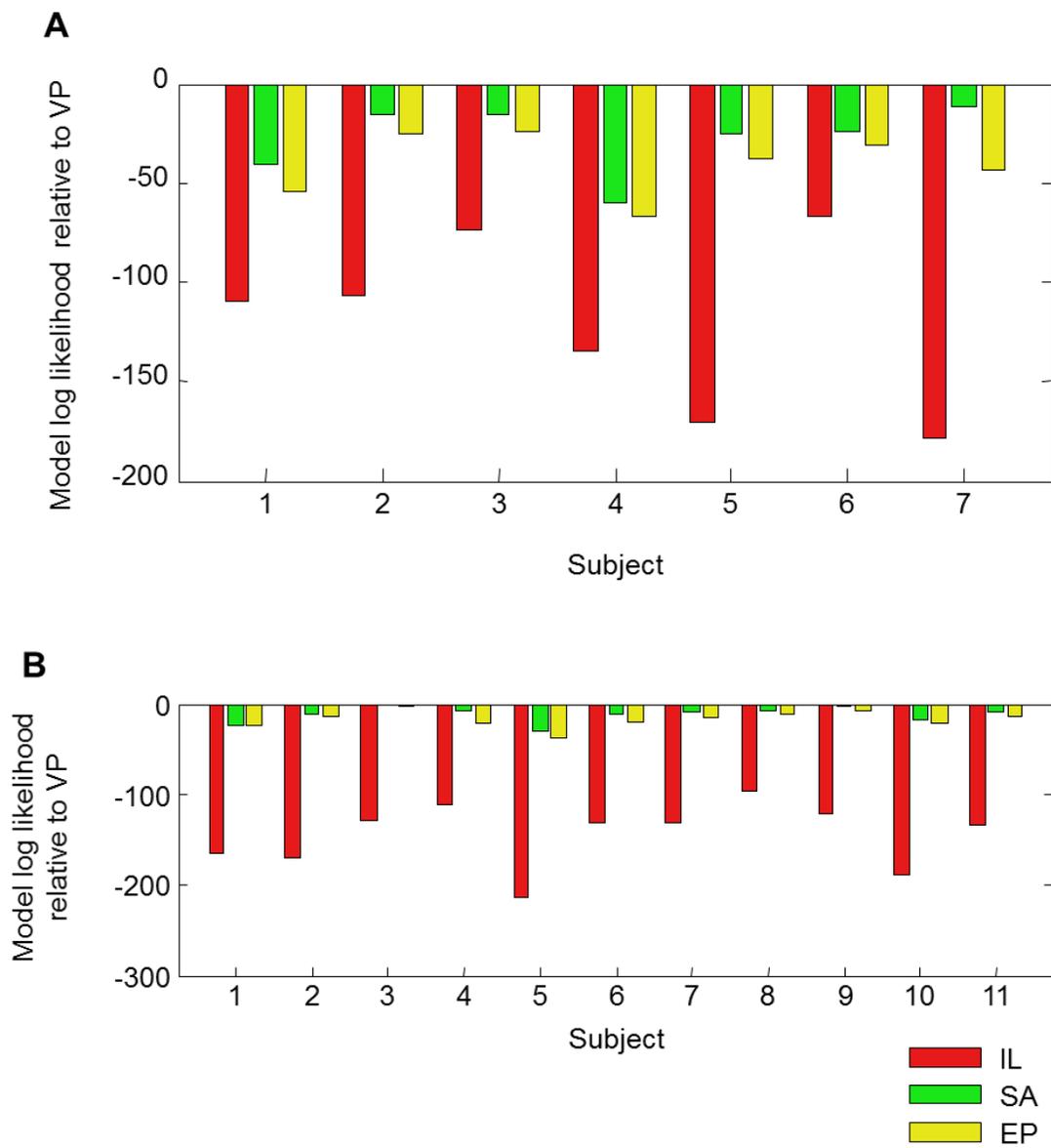


Figure 2.7 Bayesian model comparison results for individual subjects in delayed estimation. (A) Color experiment. (B) Orientation experiment.

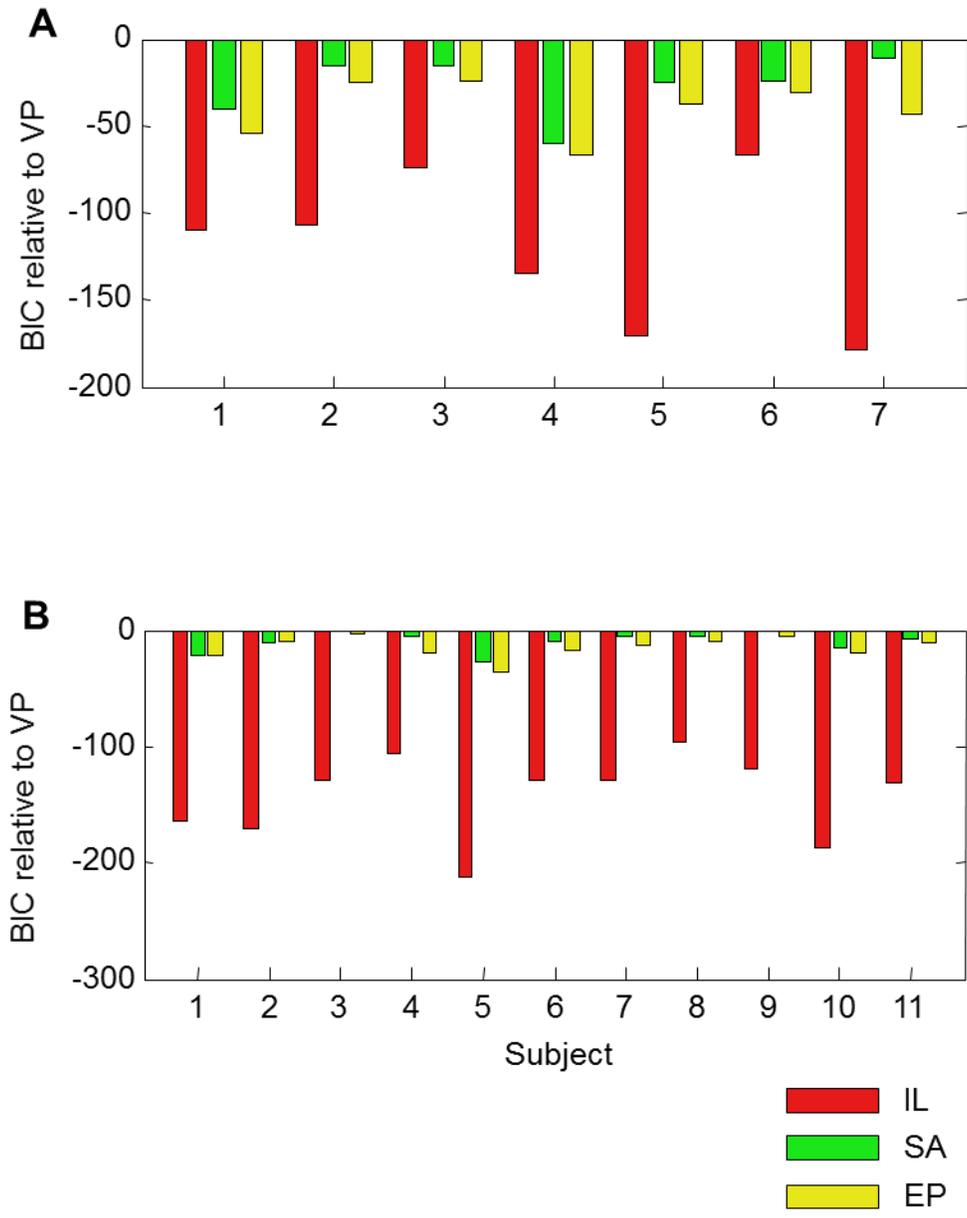


Figure 2.8 Bayesian information criterion for individual subjects in delayed estimation.

(A) Color experiment. (B) Orientation experiment.

5.3 Apparent guessing rate

To further distinguish the models, I computed an apparent guessing rate analogous to $1-w$ in delayed estimation. I did so by fitting, at each set size separately, a Bayesian-observer model with equal, fixed precision and a guessing rate to both the subject data and model-generated synthetic data. The EP model predicts an apparent guessing rate of zero. I found that subjects' apparent guessing rate was significantly higher than zero at all set sizes ($t(6) > 4.82$, $p < 0.002$ and $t(10) > 4.64$, $p < 0.001$ for Experiments 3 and 4, respectively) and increased with set size ($F(3,18) = 85.8$, $p < 0.001$ and $F(3,30) = 26.6$, $p < 0.001$, respectively). The VP model reproduces the increase of apparent guessing rate with set size more accurately than the SA model (Fig. 2.9). Like for delayed estimation, the apparent guessing rate predicted by the VP model is nonzero because items are sometimes encoded with very low precision, and this happens more frequently when set size is large.

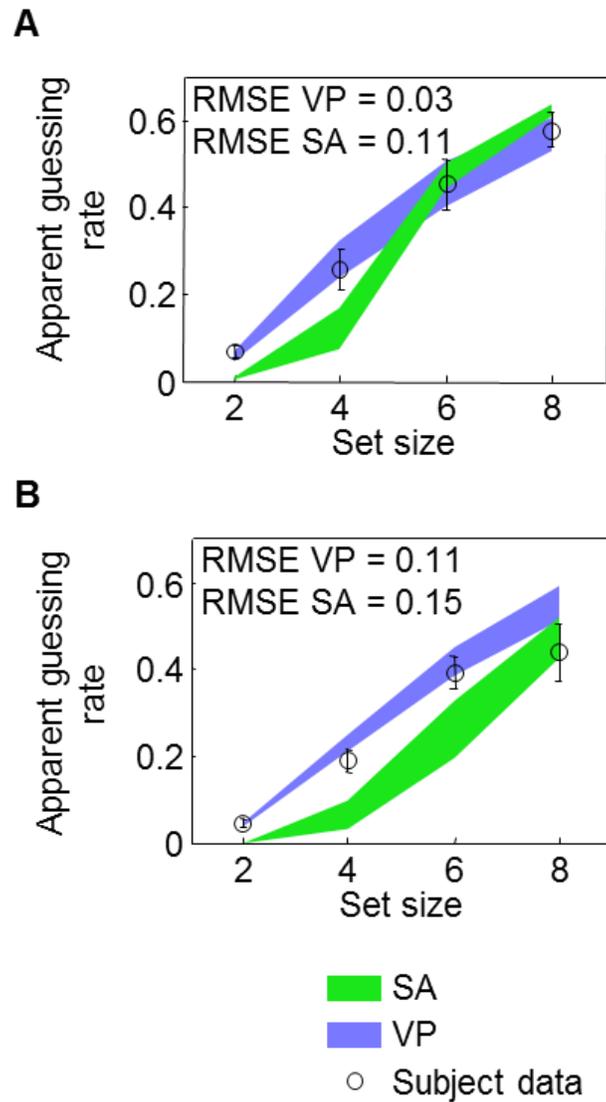


Figure 2.9 Apparent guessing rate as a function of set size. (A) Color experiment. (B) Orientation experiment.

6 Discussion

The evidence from Experiments 1 and 2 suggests that the variable-precision (VP) model better explains the limitations of VSTM than previously suggested models. Summary statistics based on average subject data shows that model prediction error is the smallest in the VP model between all models. I conducted a formal model comparison as a more rigorous way of evaluate model performance. I found that in every single subject in all experiments, the VP model performance is better than others.

Similar results were found in a delayed estimation task (54, 64). Here, subjects saw a brief display of stimuli with varying set size as in Experiments 1 and 2, and reported the estimate of a probe. The distribution of estimate errors from all set sizes were fitted. Summary statistics and Bayesian model comparison show that in each individual data from orientation and color delayed-estimation tasks, the VP model outperforms the other preexisting models that were used in this chapter. This suggests consistency of the VP model across experimental paradigms.

My results here suggest that VSTM is a continuous resource that is variable across items and trials. However, there can be other ways to implement variability in precision. In Chapter 3, I provide factorial model comparison to explore various modeling dimensions exhaustively.

Chapter 3

Factorial comparison of VSTM models

1 Introduction

In Chapter 2, I proposed the variable-precision (VP) model as an alternative to the previously existing VSTM models – the item-limit (IL), the slots-plus-averaging (SA), and the equal-precision (EP) models. Each model has a different assumption on the precision with which stimuli are encoded. In terms of number of items encoded, the IL and SA models postulate that capacity is a fixed number, but the EP and VP models assume that all items are encoded with a certain level of precision. Encoding precision and number of items are two different model factors. It is possible to combine one of the values from each factor to make a new model. For instance, combining the IL model, in which the assumption is that only a fixed number of items are encoded, with the variability in encoding precision in the VP model, gives a hybrid model in which a fixed number of items are encoded with variable precision. Moreover, it is possible to consider variability in the number of items encoded by assuming that the number of items encoded on each trial is drawn from a discrete probability distribution.

Independent model factors can be mixed to create new models. This is analogous to factorial experimental design (54), and I will call the process of comparing models created in this way factorial model comparison. Model factorization tends to generate a large number of models, and factorial model comparison is useful to efficiently rule out worse models. In addition, one can examine which model factor significantly contributes to model performance (54).

A factorial combination of the elements of the VSTM models from Chapter 2 produces a total of 14 models, including the original four. Here, I compared the predictions

of all these models on subject performance data from one-feature change localization and change detection experiments.

2 Summary of tasks

In this chapter, I study both change localization and change detection tasks. The change localization task is the same as in Chapter 2. In the change detection task, observers were shown two visual displays containing N items each, separated by a 1 second delay. Stimuli were defined by a single stimulus dimension, either orientation or color (in separate experiments). On each trial, a change could occur between the displays, with probability 0.5. When a change occurred, it occurred in one of the N items, chosen with equal probability. Observers pressed a key to report whether or not a change occurred; this is a yes-no task. Feature values are drawn independently across items from a uniform distribution over a circle in that feature space (all possible orientations for orientation; color wheel for color). If there is a change, the feature value of the changed object is also independently drawn from the same uniform distribution, and each location has the same probability of containing the change.

3 Theory and models

3.1 Model factors

In this chapter, I examine two model factors: encoding precision and number of items encoded.

3.1.1 Encoding precision at a given set size

The factor of encoding precision has four values: the infinite precision (IP), discrete chunks (SA as in the SA model), equal precision (EP), and variable precision (VP). Each value corresponds to a model from Chapter 2. The IP and SA represent the discrete-capacity view, and the EP and VP represent the continuous-resource view.

The IL model assumes perfect encoding for a fixed number of items that are randomly chosen. The chosen items are perfectly memorized, and thus, encoding precision is infinite. Thus, the value, IP, refers to the IL model. The SA model hypothesizes that memory resource comes in discrete chunks, each affording a fixed amount of precision. The EP model assumes that precision is identical among items, and the VP model describes that precision is variable among items and trials.

3.1.2 Number of items encoded

Another model factor I examine here is how many items observers can memorize or encode at a trial. This is an independent factor from encoding precision. For instance, the IL model assumes encoded items have infinite precision (IP), and the EP model assumes that precision is divided by the number of encoded items.

In the IL and SA models, the number of items encoded was a fixed capacity K . However, in the EP and VP models, all items are encoded. Inspired by the idea of variability in precision in the VP model, I allow for the possibility that K varies from trial to trial. I consider two forms of variability; at each trial, K is drawn from a uniform distribution; at each trial, K is drawn from a Poisson distribution. This results in four values in the factor of number of items encoded; fixed, all, variable-uniform, and variable-Poisson.

3.2 Factorial combination of the modeling dimensions

I considered all possible models based on factorial combination of these two model factors: encoding precision and number of items encoded. Each of the four factorss has four values, which generates a total of $4 \cdot 4 = 16$ models. However, in the discrete-capacity models (the IL and SA), observers cannot encode all items because in this case, they exhibit perfect memory for all items regardless of set size.

3.2.1 *Models with infinite K*

When the precision is infinite (the IL model) or discrete chunks (the SA model), infinite capacity (or number of slots) always gives perfect performance regardless of set size because all items are perfectly encoded. Thus, I will not consider the IL model or the SA model with infinite K . The EP and VP models assume that all items are encoded. Therefore, there are two models with infinite K .

3.2.2 *Models with fixed K*

The IL and SA models assume that observer have a fixed capacity K . Equal-precision value can be combined with fixed- K . In this model, observers encode K items with equal precision. Compare to the EP model, which assumes all items are encoded, the probability of a correct response now depends on the relationship between K and N because the changed item, namely “target” can be among the encoded items or not depending on the relationship between K and N . There can be two cases. First, if $K \geq N$, $\kappa = \Phi(J_1 N^\alpha)$ (see Eq. 2.6 in Section 3.3.1 in Chapter 2). When $K < N$, there are two possibilities. The target may exist among the selected K items with the probability of K/N , and $\kappa = \Phi(J_1 K^\alpha)$. If not,

observers have to make a guess about the location of the target. Similarly, in the combination of the VP with fixed K , K items are encoded with variable precision at each trial. If $K \geq N$, the precision of each item is drawn from a Gamma distribution with a mean $\bar{J} = \bar{J}_1 N^\alpha$, and if $K < N$, $\bar{J} = \bar{J}_1 K^\alpha$ (see Eq. (2.10) in Section 3.3.3. in Chapter 2).

The EP model with fixed K , and the VP model with fixed K , have an additional capacity parameter, the fixed capacity K , compared to the EP and VP models that assume all items are encoded, respectively.

3.2.3 *Models with Poisson-distributed K*

One way of implementing variability in number of items encoded is to assume that capacity K is drawn from a discrete probability distribution (54). One assumption is that number of items encoded is drawn from a Poisson distribution.

In the models with Poisson-distributed K , at each trial, capacity K is drawn from a Poisson distribution with a mean K_{mean} . The predictions of the models with Poisson-distributed K are computed by taking a weighted average for each K . I computed the probability of each K drawn from a Poisson distribution with a mean K_{mean} . K can be any natural number, and thus, I made a cut-off when $K > 25$ because probability mass after $K > 25$ is ignorable. This probability was computed for $K = 1, \dots, 8$, and I considered any K larger than 8 has the same probability. Response probability was weighted by this precomputed probability.

In terms of model parameters, the IL and SA models with Poisson-distributed K now have a parameter of K_{mean} instead of fixed K . The EP and VP models with Poisson-

distributed K have K_{mean} as a parameter in addition to the mean and scale parameters of the Gamma distribution. This results in four models with Poisson-distributed K .

3.2.4 Models with uniformly distributed K

Another way of implementing variability in the number of items encoded is to assume that K is drawn from a uniform distribution. In the models with uniformly distributed K , at each trial, K is an integer and is drawn from a discrete uniform distribution on $[0, K_{\text{max}}]$. K_{max} is an integer that ranges from 1 to 8, and remains constant during all trials. Predictions of the models with uniformly distributed K is computed by averaging predictions of the models with fixed K across all values of K between 0 and K_{max} . This is similar to applying probability weights to response probability in the case of Poisson-distributed K .

In terms of model parameters, the IL and SA models with uniformly distributed K now have a parameter K_{max} instead of a fixed K . The EP and VP models with uniformly distributed K have K_{max} as parameter in addition to the mean and scale parameters of the Gamma distribution. This results in four models with uniformly distributed K . This results in four models with uniformly distributed K .

In total, this factorial combination gives rise to 14 models (Fig. 1); four existing and ten new models.

3.3 Nomenclature

I use a three-letter code to give a name to each model. The first two letters indicate the model's assumption about encoding precision of a model; IP: infinite encoding precision

(as in the IL model), SA: discrete chunks of encoding precision (as in the slots-plus-averaging model), EP: equal precision, and VP: variable precision.

The last letter indicates assumptions about the number of items encoded; A: all items are encoded, F: a fixed number of items are encoded, P: the number of items encoded at each trial is drawn from a Poisson distribution, and U: the number of items encoded at each trial is drawn from a uniform distribution.

For example, the VPF model is the model in which a fixed number of items are encoded (F) at each trial, and among these items, encoding precision is variable. The four existing models discussed in Chapter 2 can now be denoted by IPF, SAF, EPA, and VPA.

3.4 Models

All 14 models that are used in this chapter are shown in Fig. 3.1. Table 3.1 summarizes the parameters of each model.

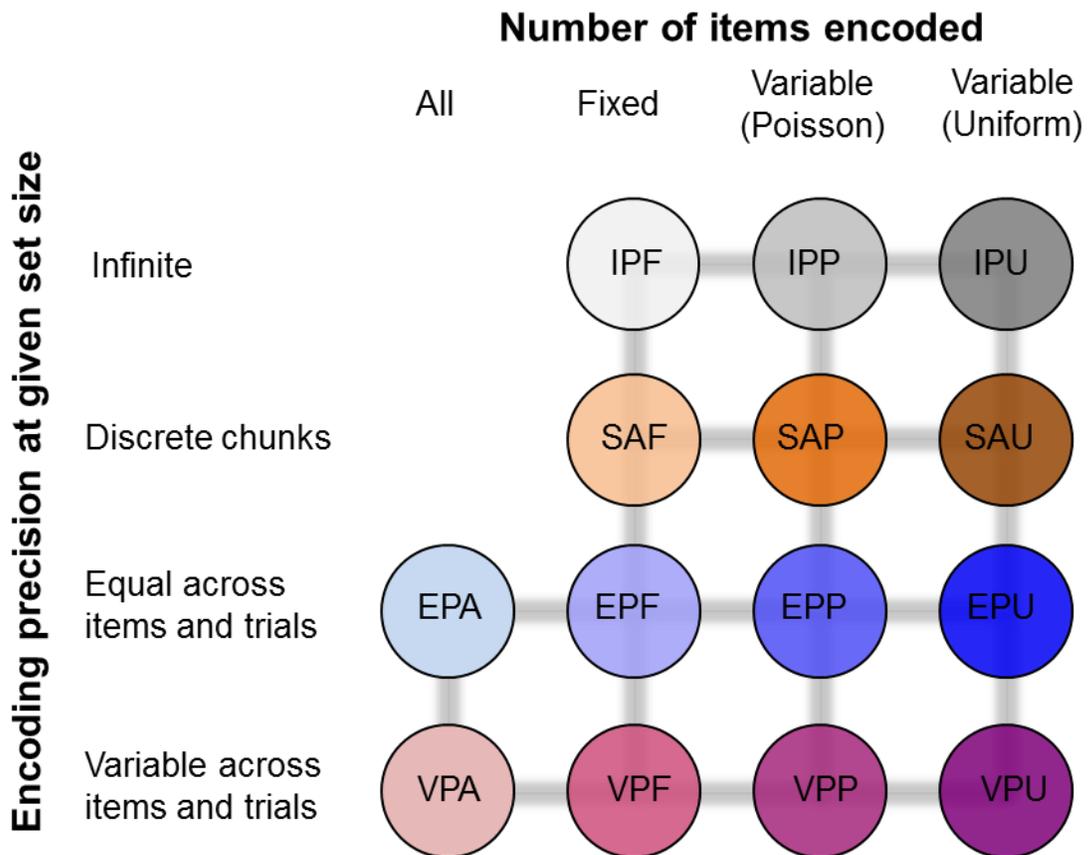


Figure 3. 1 Models of VSTM. Based on two modeling criteria (Encoding precision and number of items encoded), 14 combination of models (circles) are suggested for exhaustive model comparison. The basic VSTM models from Chapter 2 are IPF, SAF, EPA and VPA

Models	Encoding precision	Number of items encoded	Model parameters
IPF	Infinite	Fixed	K, ε
IPP	Infinite	Variable (Poisson)	$K_{\text{mean}}, \varepsilon$
IPU	Infinite	Variable (Uniform)	$K_{\text{max}}, \varepsilon$
SAF	Discrete chunks	Fixed	K, J_1
SAP	Discrete chunks	Variable (Poisson)	K_{mean}, J_1
SAU	Discrete chunks	Variable (Uniform)	K_{max}, J_1
EPA	Equal	All	J_1, α
EPF	Equal	Fixed	K, J_1, α
EPP	Equal	Variable (Poisson)	$K_{\text{mean}}, J_1, \alpha$
EPU	Equal	Variable (Uniform)	$K_{\text{max}}, J_1, \alpha$
VPA	Variable	All	\bar{J}_1, α, τ
VPF	Variable	Fixed	$K, \bar{J}_1, \alpha, \tau$
VPP	Variable	Variable (Poisson)	$K_{\text{mean}}, \bar{J}_1, \alpha, \tau$
VPA	Variable	Variable (Uniform)	$K_{\text{max}}, \bar{J}_1, \alpha, \tau$

Table 3. 1 Summary of the VSTM models tested in Chapter 3.

3.5 Bayesian inference in change detection

The generative model of a change detection task is illustrated in Fig. 3.1. Change occurrence, denoted by C , is a binary variable (0: no change; 1: change). Change detection is similar to change localization (Fig. 3.2) except that now Bayesian inference is used to compute the probability of change occurrence, C , given measurements, \mathbf{x} and \mathbf{y} . The decision variable, d , is the ratio of the probability of change given measurements and the probability of no-change given measurements, which is the posterior ratio of change occurrence. If d is larger than 1, it means the observer's belief in change occurring during a trial is larger than the belief in no-change during the same trial. Using Bayes' theorem, the posterior ratio becomes the product of likelihood ratio and prior ratio. The likelihood ratio is

$$\frac{1}{2\pi N} \sum_{i=1}^N \frac{\iint p(x_i | \theta_i) p(y_i | \varphi_i = \theta_i + \Delta) d\theta_i d\Delta}{\int p(x_i | \theta_i) p(y_i | \varphi_i)}$$

and the prior ratio is $\frac{P_{\text{change}}}{1 - P_{\text{change}}}$, where p_{change} is an observer's prior belief of change

occurrence, which assumed to be 0.5 in my experiment. Therefore, the prior ratio is 1, and thus the decision variable becomes

$$d = \frac{1}{2\pi N} \sum_{i=1}^N \frac{I_0(\kappa_i)^2}{I_0(\kappa_i \sqrt{2 + 2\cos(x_i - y_i)})}$$

For a given parameter combination of a model, and a given change magnitude Δ , I calculated the probability of a correct response to compute model prediction.

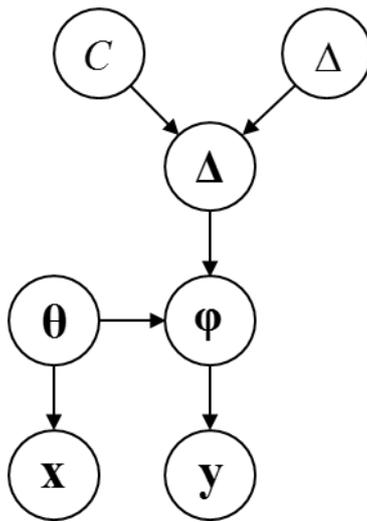


Figure 3.2 Generative model for the change detection task. C : change occurrence (0: no change; 1: change); Δ : magnitude of change; Δ : vector of change magnitudes at all locations; θ and φ : vectors of stimuli in the first and second displays, respectively; x and y : vectors of measurements in the first and second displays, respectively.

4 Methods

4.1 Psychophysics

4.1.1 *Change localization*

I used the same set of data collected in Chapter 2 (see details in Section 4.1 in Chapter 2).

4.1.2 *Change detection*

Stimuli were the same as those used in the change localization experiments. In this change detection task, subjects first saw a brief display (100 ms). After a delay (1000 ms), subjects saw another brief display (100 ms). For the half of the trials, one of the items on the second display had changed (the “change” trial). The change magnitude was drawn from a uniform distribution for the change trials. After the second disappeared, subjects decided whether this trial was a change-trial or no-change-trial. In the orientation experiment, set size varied: 2, 4, 6 and 8 as in the change localization experiments. In the color experiment, set size was: 1, 2, 4 and 8. For both experiments, each subject completed 1800 trials. Ten subjects participated in an orientation change detection experiment, and seven subjects in a color change detection experiment. The trial procedures of both change detection experiments are shown in Fig. 3.3.

4.2 Model comparison

As in Chapter 2, in addition to comparing summary statistics, I used Bayesian model comparison to compute model performance for each experiment (see details in Section 4.2 in Chapter 2).

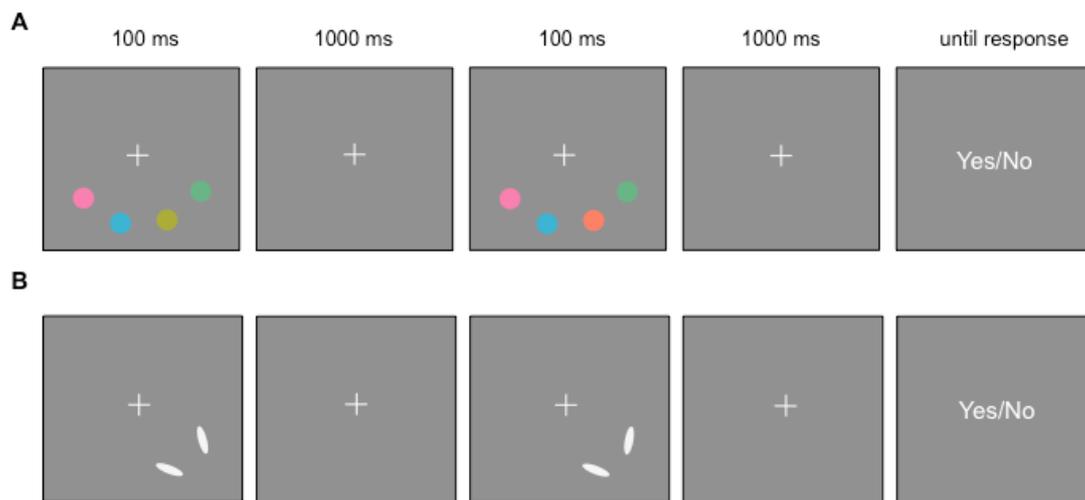


Figure 3.3 Trial procedure of change detection. (A) Color experiment. (B) Orientation experiment.

4.3 Model recovery analysis

To validate the implementation of each model, I conducted model recovery analysis. One can generate a synthetic dataset based on a model's assumption. Let's call this model a parent model. When the dataset is fitted to multiple different models including the parent model, if one compares the predictions of these models, the parent model should give the best prediction among others for this specific synthetic dataset.

To implement this model recovery analysis, for each model, five synthetic change detection datasets are generated by a random combination of parameters of the model. Based on each synthetic dataset, I computed the log likelihood for all models to each synthetic dataset and compared the mean model likelihoods of synthetic datasets.

This model recovery analysis (Fig. 3.4) shows that implementation of each model can correctly recover the parent model based on its synthetic dataset.

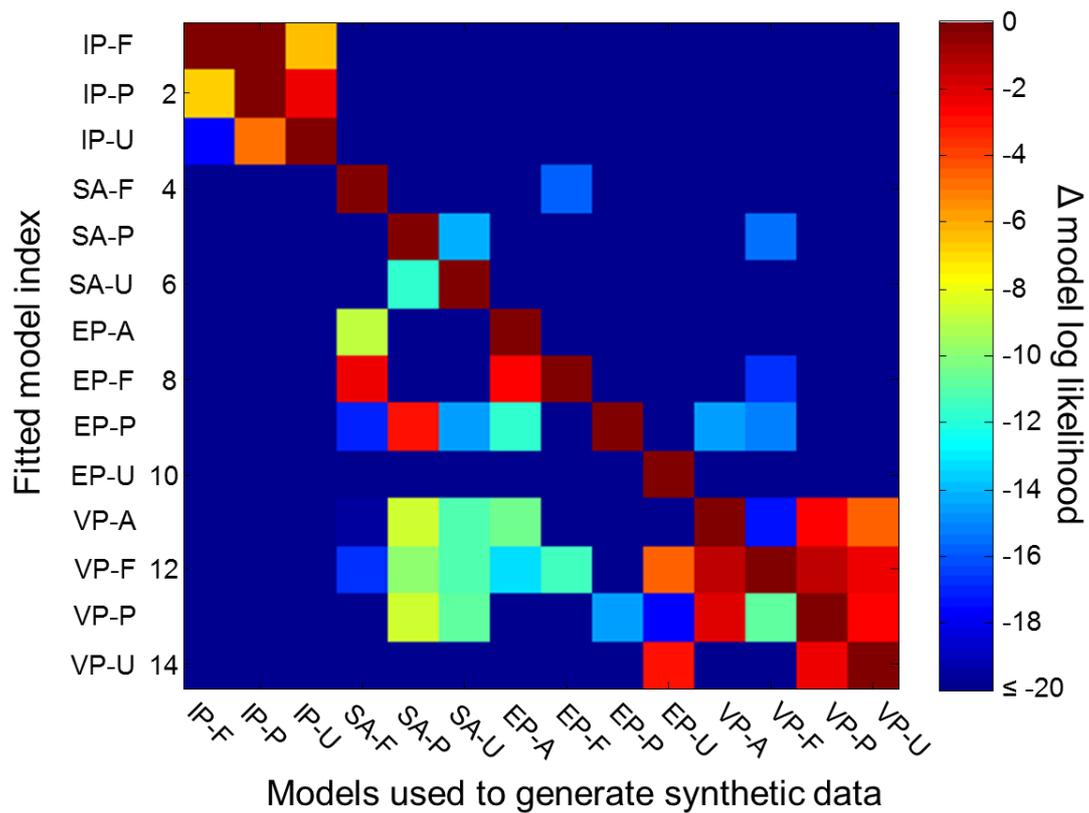


Figure 3.4 Model recovery analysis. We tested how well synthetic data sets generated from each model (columns) were fitted by each model (rows). The color in a cell indicates log likelihood (log Bayes factor) of all fitted models relative to the winning model for the corresponding synthetic data. The darkest red on the diagonal means that the model used to generate the data was found to be most likely.

5 Results

5.1 Model comparison

I computed the model log likelihoods (log Bayes factor) of all 14 models for all subjects for each experiment (orientation change localization, color change localization, orientation change detection, and color change detection). For each experiment, I found the winning model, which has the largest mean log likelihood across subjects. Then, for each subject, I subtracted the log likelihood of the winning model from the log likelihood of each model. Then I computed the mean and s.e.m. of the differences in log likelihood. Based on Jeffreys (65), I consider that models, in which log likelihood difference is larger than $\log 30$ compared to the winning model, are substantially worse than the winning model.

5.1.1 *Change localization*

In the orientation experiment, the winning model was the SAP model (Fig. 3.5A), where memory precision is defined as discrete chunks and capacity for each trial is drawn from a Poisson distribution. However, these following models have log likelihood differences relative to the winning SAP model, less than $\log 30$: VPA(-1.09 ± 0.50), VPF(-1.17 ± 0.71), VPP(-1.18 ± 0.56), EPU(-1.56 ± 0.67), SAU(-1.81 ± 0.91), and EPP(-2.6 ± 1.2). Thus, I consider the predictions from these six models indistinguishable from the winning model. Model fits to subject performance in the orientation experiment of each model are shown in Fig. 3.6.

In the color experiment, the winning model was the VPP model (Fig. 3.5B), where memory precision is variable across items and trials, and capacity for each trial is Poisson-distributed. Difference in log likelihood of the five following models were

less than $\log 30$: VPU(-0.84±0.80), SAP(-0.9±1.0), EPP(-0.93±0.94), VPF(-0.97±0.35), and VPA(-2.01±0.62). This shows that these five models are indistinguishable from the winning VP-P model. The model fits to subject performance in the color experiment of each model are shown in Fig. 3.7. The differences in log likelihood of all models for the orientation and color change localization experiments are shown in Table 3.2.

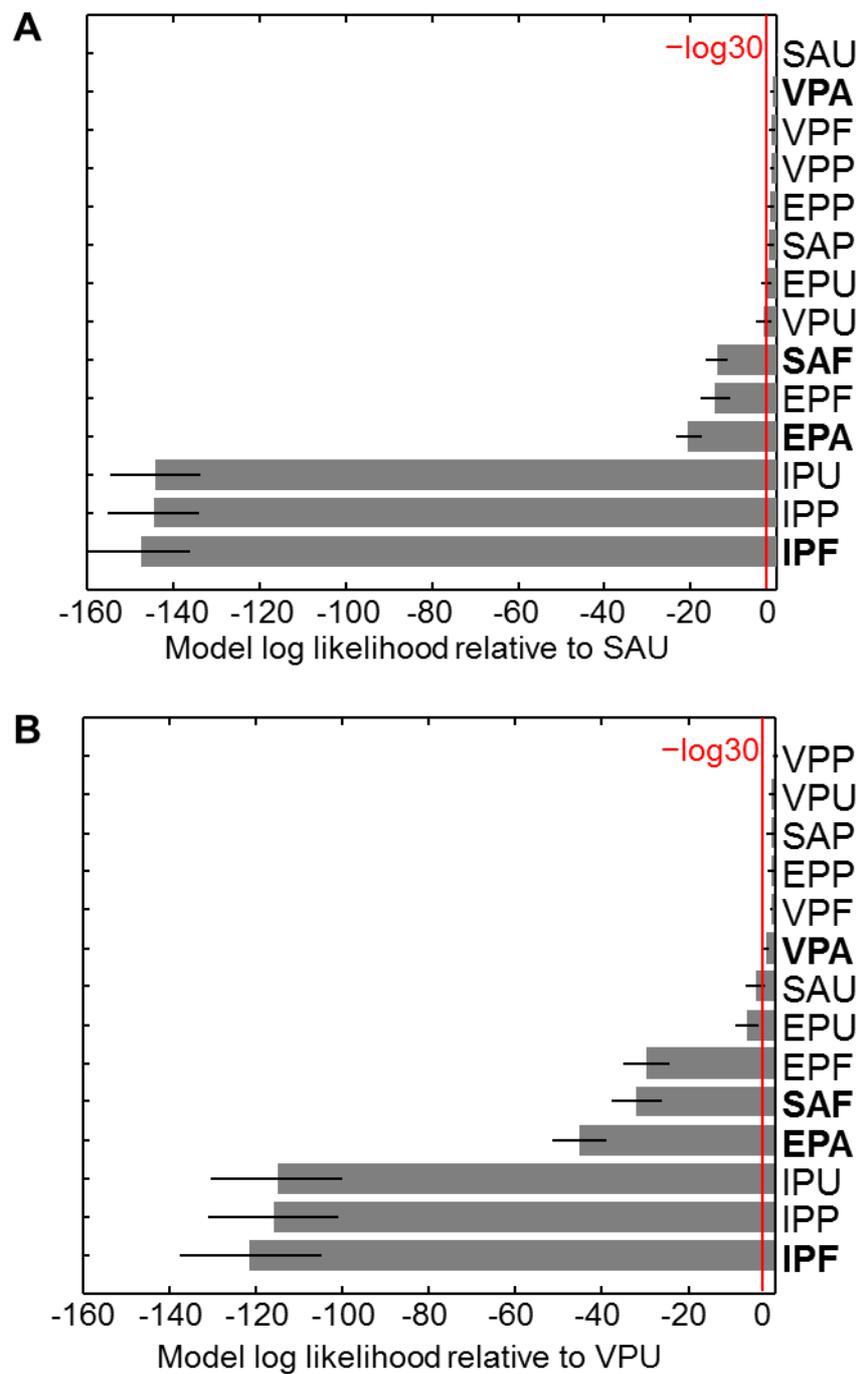


Figure 3.5 Relative log likelihood of models to the winning model in descending order for the orientation (the SAP) and color (the VPP) change localization experiments. Models with bold letters are the ones used in Chapter 2.

Orientation (winner: SAP)		Color (winner: VPP)	
Models	mean±s.e.m.	Models	mean±s.e.m.
*VPA	-1.09±0.50	VPU	-0.84±0.80
VPF	-1.17±0.71	SAP	-0.9±1.0
VPP	-1.18±0.56	EPP	-0.93±0.94
EPP	-1.56±0.67	VPF	-0.97±0.35
SAU	-1.81±0.91	*VPA	-2.01±0.62
EPU	-2.6±1.2	SAU	-4.6±2.1
VPU	-3.2±1.6	EPU	-6.6±2.7
*SAF	-14.0±2.5	EPF	-29.8±5.4
EPF	-14.3±3.4	*SAF	-32.0±5.9
*EPA	-20.6±3.0	*EPA	-45.2±6.4
IPU	-144±10	IPU	-115±15
IPP	-145±11	IPP	-116±15
*IPF	-147±11	*IPF	-121±16

Table 3.2 Mean and s.e.m. of relative log likelihood of each model to the winning model for the orientation (the SAP) and color (the VPP) change localization experiments. Models with bold letters give indistinguishable model prediction compared to the winning model. Models that were used in Chapter 2 are marked with asterisks.

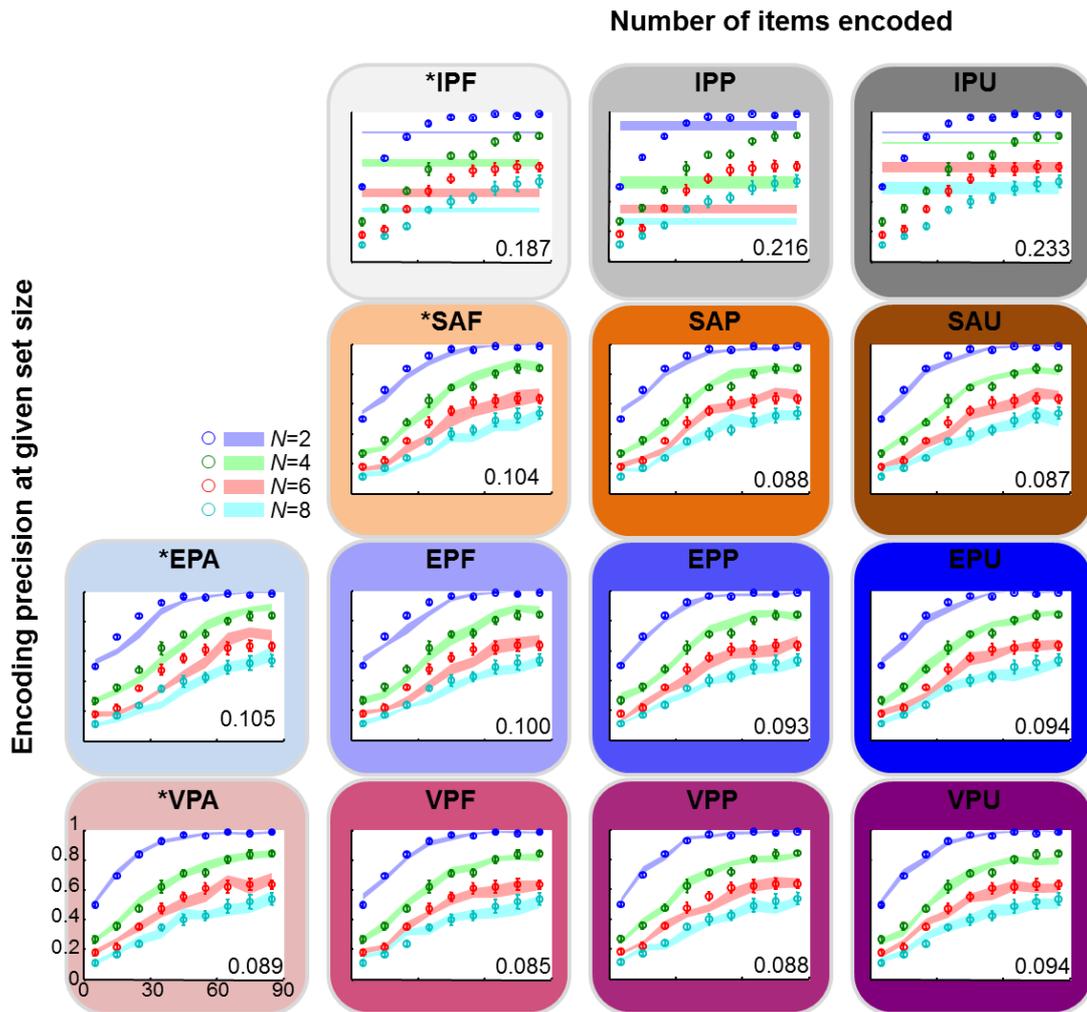


Figure 3.6 Model fits to subject performance from the orientation change localization experiment. The names of the models tested in Chapter are marked with *. The root-mean-square error (RMSE) is shown for each model.

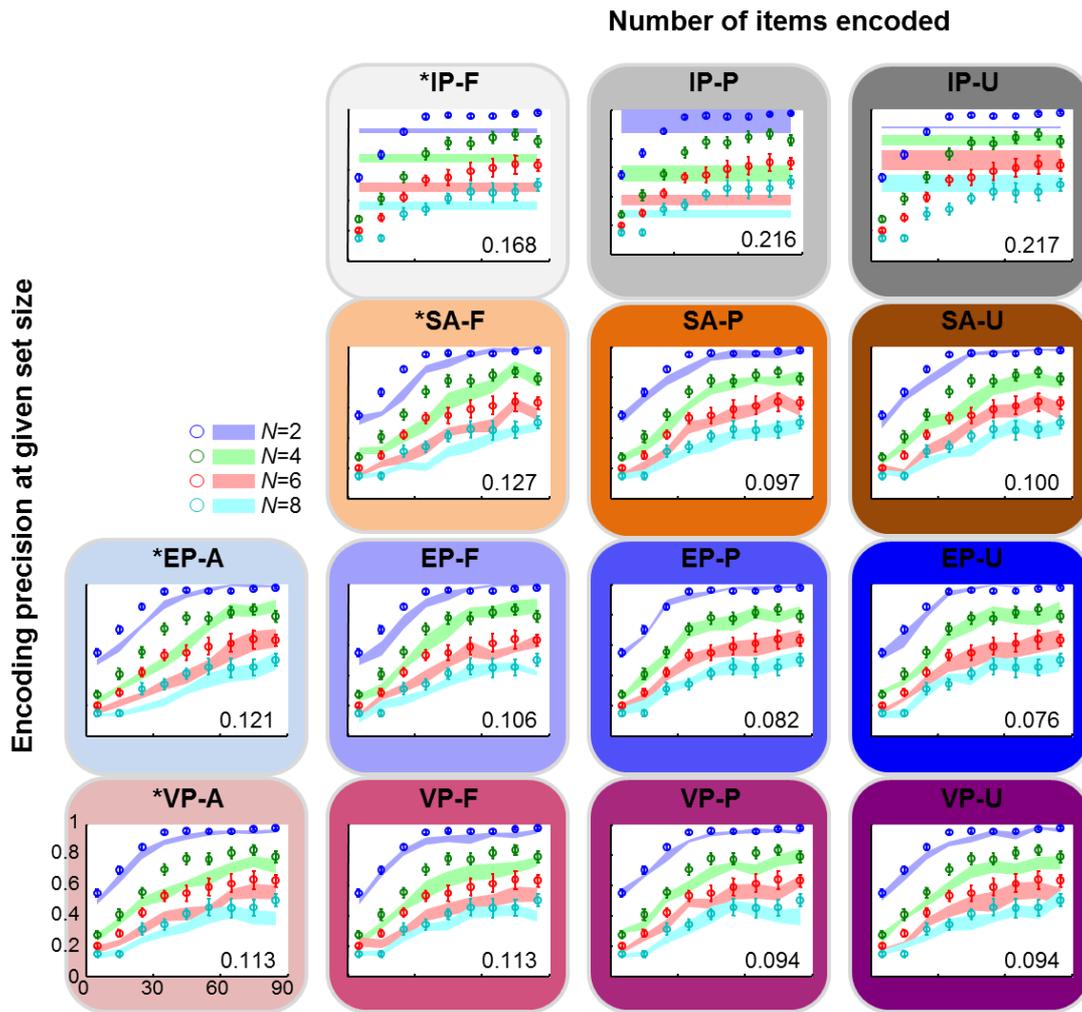


Figure 3.7 Model fits to subject performance from the color change localization experiment.

5.1.2 *Change detection*

In the orientation experiment, the winning model was the VP-F model (Fig. 3.8A), according to which memory precision is variable across items and trials and capacity is fixed. The VPA and VPP models have log likelihood difference less than log 30: VPA(-1.25±0.62) and VPP(-1.26±0.37). Thus, the predictions from the two models are indistinguishable from the winning model. Model fits to subject performance in the orientation experiment of each model are shown in Fig. 3.9.

In the color experiment, the winning model was the VPF model (Fig. 3.8B), just like it was in the orientation experiment. The VPP model was not distinguishable from the winning model (difference in log likelihood: -2.6±1.2). Model fits to subject performance in the color experiment of each model are shown in Fig. 3.10. The differences in log likelihood of all models for the orientation and color change detection experiments are shown in Table 3.3.

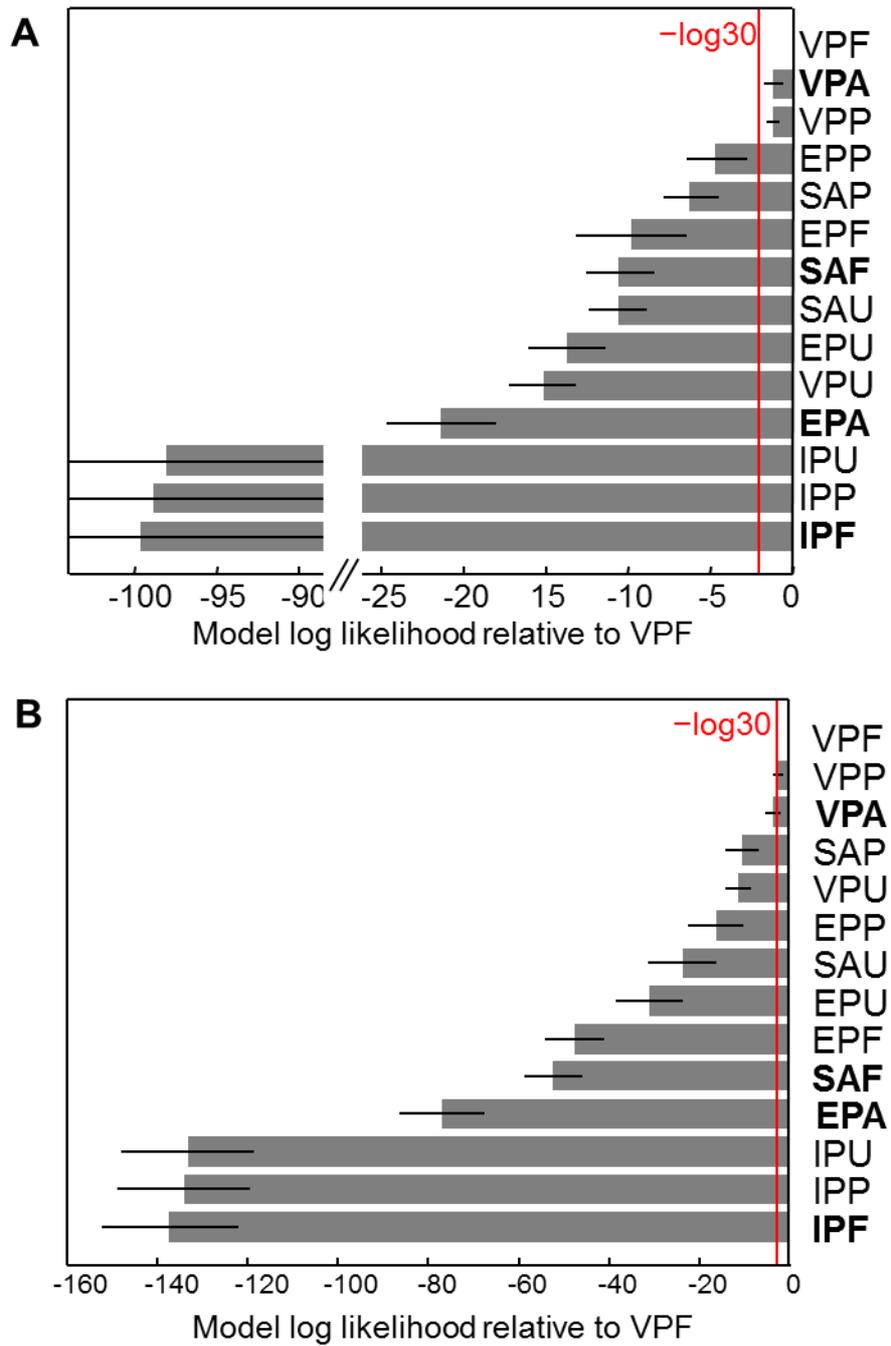


Figure 3.8 Relative log likelihood of models to the winning model (the VP-F) in descending order for the orientation and color change detection experiments. Models with bold letters are the ones used in Chapter 2.

Orientation (winner: VPF)		Color (winner: VPF)	
Models	mean±s.e.m.	Models	mean±s.e.m.
*VPA	-1.25±0.62	VPP	-2.6±1.2
VPP	-1.26±0.37	*VPA	-3.7±1.7
EPP	-4.7±1.9	SAP	-10.5±3.7
SAP	-6.3±1.7	VPU	-11.4±2.8
EPF	-9.9±3.4	EPP	-16.4±6.2
*SAF	-10.6±2.0	SAU	-23.9±7.4
SAU	-10.7±1.8	EPU	-31.0±7.4
EPU	-13.7±2.3	EPF	-47.6±6.6
VPU	-15.2±2.0	*SAF	-52.4±6.5
*EPA	-21.5±2.0	*EPA	-77.1±9.4
IPU	-98±11	IPU	-133±15
IPP	-99±11	IPP	-134±15
*IPF	-99±11	*IPF	-137±15

Table 3.3. Mean and s.e.m. of relative log likelihood of each model to the winning model (the VPF) for the orientation and color change detection experiments. Models with bold letters give indistinguishable model prediction compared to the winning model. Models that were used in Chapter 2 are marked with asterisks.

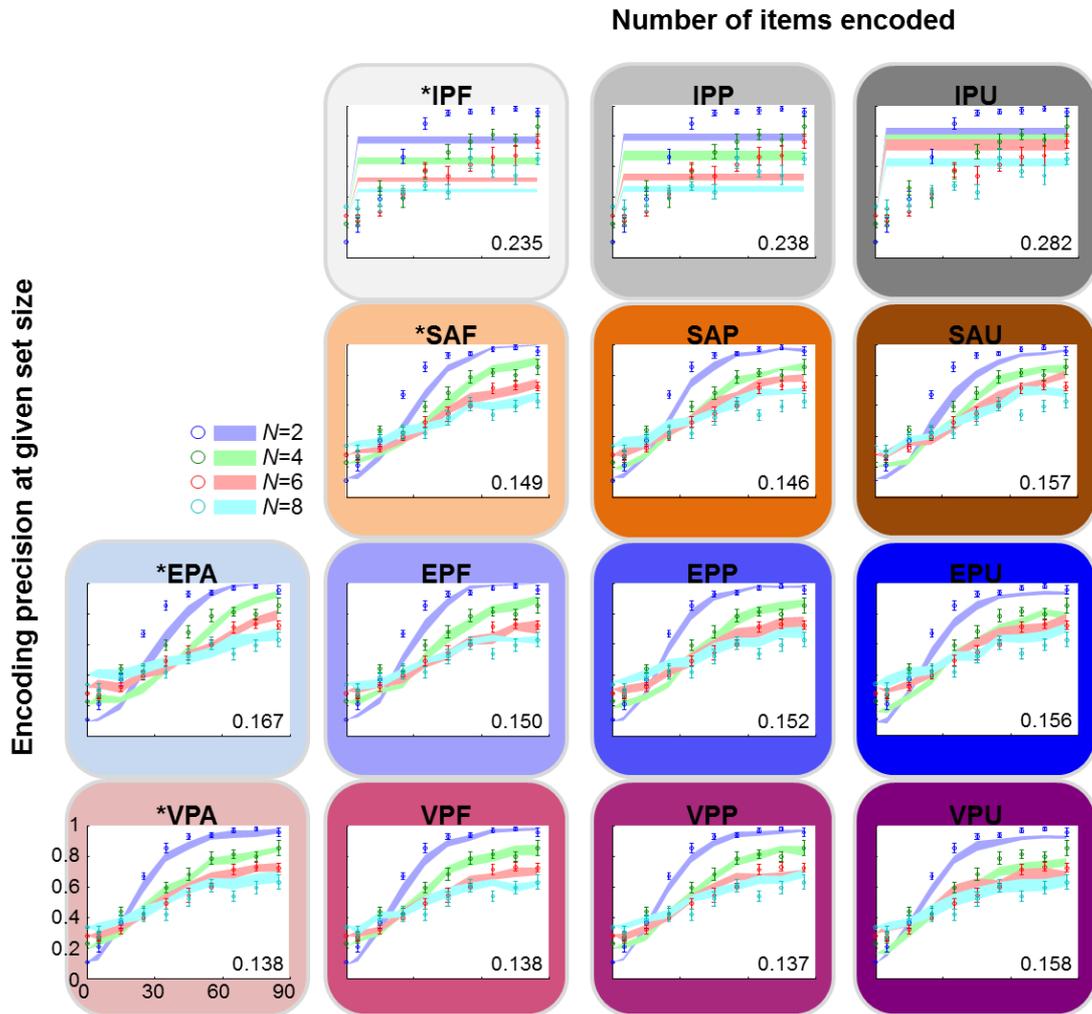


Figure 3.9 Model fits to subject performance from the orientation change detection experiment.

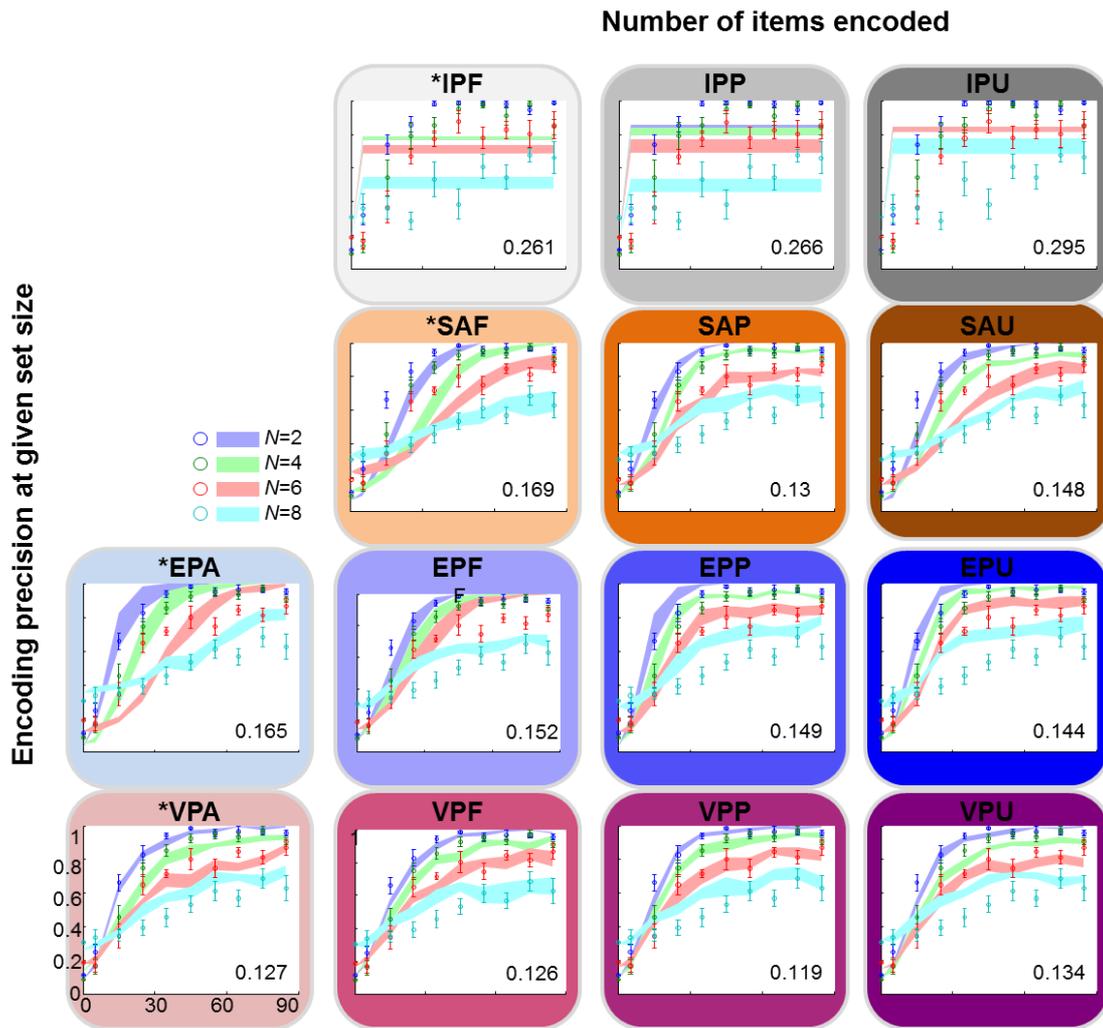


Figure 3.10 Model fits to subject performance from the color change detection experiment.

5.2 Statistical dependence across tasks and features

To examine whether the factorial model comparison results are invariance across tasks and features, I computed Spearman's rank correlation coefficient, ρ . Between the two tasks, correlation coefficient was positively correlated with significance in both features ($\rho_{\text{CL-CD,orientation}} = 0.82, p < 0.001, \rho_{\text{CL-CD,color}} = 0.91, p < 0.001$). Between the two features, correlation coefficient was also positively correlated with significance in both tasks ($\rho_{\text{orientation-color,CL}} = 0.75, p = 0.003, \rho_{\text{orientation-color,CD}} = 0.89, p < 0.001$).

6 Discussion

In this chapter, I used a factorial model comparison method to explore two major model factors from VSTM literature, encoding precision and number of items encoded. I compared 14 models, which include four models already discussed in Chapter 2; the IL, SA, EP, and VP models. I used change detection paradigm in addition to change localization.

I found that between the four models that I examined in the Chapter 2, the variable precision model (the VPA) can provide better prediction on subject behavior than other three models (the IPF, SAF, and EPA in this chapter).

In all experiments, at least one winning model has variability in encoding precision. In the change localization task, more than half of the VP variants are the winning models (orientation: VPA, VPF, and VPP; color: all four VP variants). The change detection experiments show that all winning models are one of the VP variants (orientation: VPF and VPA; color: VPF). This suggests that variability in encoding precision majorly contributes to explaining performance in the VSTM tasks.

However, models in which encoding precision is discrete generally perform poorly. All IP variants never made the cutoff. In the case of SA variants, the SAU (orientation) and SAP (orientation and color) are considered as winning models only in the change localization experiments. In the change localization experiments, non-VP models made the cutoff. However, these non-VP models were either with Poisson-distributed K (the SAP and EPP) or uniformly distributed K (the SAU). This implies that any form of variability (either encoding precision or number of items encoded) might be needed for a better VSTM model.

The rank of model performance was consistent between the two tasks within orientation and color respectively, and between the two features within change localization and change detection respectively. This shows that the results from model comparison on data are highly consistent across different tasks, suggesting that the true nature of VSTM is invariant across tasks.

The results from the current and previous chapters show that the continuous-resource view, especially the variable-precision model, provides better insight on human VSTM. In the following two chapters, I apply the continuous-resource view to address VSTM limitations in more naturalistic stimuli.

Chapter 4

VSTM for two-feature objects (1):

memory allocation to relevant and irrelevant features

1 Introduction

To examine limitations of VSTM, the previous two chapters used simple stimuli. However, these stimuli are far from complex and rich natural stimuli. As a stepping stone to more naturalistic stimuli, I used objects with multiple features in Chapters 4 and 5, to further investigate limitations of VSTM.

Most of the time, studies on VSTM for multi-feature objects have been framed as investigations on whether the basic unit of VSTM is an object or a feature. In this object-versus-feature framework, if the basic unit is an object, memory is limited by the number of objects, and if it is a feature, by the number of features. However, this framework is still based on the discrete-capacity view of VSTM limitations. Based on this view, two questions have been widely asked regarding the unit of VSTM storage for multi-feature items:

H1. If VSTM is object-based, then the memory of a feature should not suffer from the addition of a second feature to the same object. Indeed, whether an object has orientation, color, or both does not seem to affect performance in a change detection task (28, 31, 33, 66), ruling out that VSTM stores a fixed total number of features (summed across all objects). Luck and Vogel (28) found that subject performance was similar whether an object has only one feature or four features (orientation, color, spatial distance, and size). However, this finding is compatible both with a VSTM system that stores (encodes) K objects regardless of their number of features (28, 31) and with one in which each feature has a separate capacity of K (33, 66). The former does not allow a feature to have its own capacity, and the latter argues that according to the former, object capacity, which is the sum of all feature capacities, is not fixed but increases with the

number of features in an object. The latter view is related to the feature-integration theory of attention (67), which suggests that features are automatically registered at the early stage of visual perception, and objects are only identified at later stages by focused attention. This theory suggests that each feature that is encoded in parallel can be “glued” together only if attention integrates these features into an object.

H2. If VSTM is object-based, then encoding the task-relevant feature of an object should automatically cause irrelevant features of that object to be encoded as well. This hypothesis has only been tested by examining whether the addition of an irrelevant feature decreases performance. The finding that it does not (31, 50, 68, 69) (but see 8, 9) has been interpreted as evidence for the irrelevant feature not being encoded (31, 50). Vogel et al. (31) found that simple addition of an irrelevant feature to stimuli (e.g., from gray bars to colored bars) did not affect subject performance. Luria and Vogel (50) found similar results; addition of irrelevant colors to shape stimuli did not impair performance. However, this interpretation relies on two implicit assumptions: that the irrelevant feature does not have its own capacity (similar to H1), and that if irrelevant feature were encoded, it would be ignored during decision-making (a problem that to our knowledge has not previously been noted).

To answer these questions, earlier studies have used the same paradigm: change detection with “highly discriminable” stimuli, such as primary and secondary colors (28, 31–33, 66, 68–72). Such stimuli were chosen with the intention of avoiding the various sources of noise corrupting VSTM. The idea is that if the magnitude of the change is large compared to the noise, then one can regard the internal representation

of the remembered stimuli as noiseless. However, as Chapter 2 and 3 suggested, memories are not encoded perfectly and many recent studies support this view (35, 39, 40, 42, 51, 54, 64); memories are noisy and that a complete account of VSTM should take noise into account.

In change detection and similar paradigms, a noise-based view has two main consequences. First, an object is no longer in a binary “encoded” versus “not encoded” state; instead, it is encoded with a certain level of precision (which can be zero) (54). Second, retrieval, including decision-making, becomes an active process of probabilistic inference (73, 74): the observer decides which hypothesis – change or no change – is better supported by the noisy stimulus measurements from both displays.

If memories are noisy, then the previous views of object-based VSTM, as captured in hypotheses H1 and H2, also have to be modified (41, 75). Here, I simultaneously address these hypotheses using change localization with a variable magnitude of change (36, 40, 58, 64, 73, 74). Varying the magnitude of change allows a precise description of the role of memory noise: more noise means a shallower psychometric curve over change magnitude. Using “highly discriminable” stimuli, like the classic studies do, would amount to measuring only one point on this psychometric curve – and it is even unclear which point. Concluding that performance at that point is the same in two conditions leaves open the possibility that it is different at other points. Therefore, measuring full psychometric curves over change magnitude allows for stronger conclusions on whether performance is unaffected by our manipulations than using “highly discriminable” stimuli across a change.

To translate the hypotheses to the setting of noisy memories, it is convenient to introduce the concept of resource (51). More memory resource to a stimulus implies

that the measurements of that stimulus have lower noise or higher precision. The questions corresponding to H1 and H2 are then as follows.

(Q1) Resource allocation among features. In the resource view, there is no clear analog of the classic notion of object-based VSTM that VSTM “stores K objects regardless of their number of features” (28, 31), because as soon as one accepts that storage is not perfect but happens with finite precision, one has to specify this precision for every feature. However, one can still test the equivalent of the hypothesis by Wheeler, Treisman, Olson, and Jiang that each feature has a separate capacity (33, 66) – it becomes the question “is memory resource shared among features of an object, or does each feature have its own resource?”

(Q2) Encoding and role in decision-making of irrelevant features. Does an irrelevant feature receive resource? If the answer is yes, there is a follow-up question: in the process of deciding which location contains the change based on noisy memories (36, 39, 55, 73), does the irrelevant feature get ignored? Distinguishing “not encoded” from “encoded but ignored during decision-making” addresses a confound already present in the noiseless view of VSTM (H2 above).

2 Theory and models

Combining the possible answers to Q1 and Q2 results in six models (Fig. 4.1). In Models 1 to 3, orientation and color share memory resource, whereas in Models 4 to 6, orientation and color have independent pools of resources. In Models 1 and 4, an irrelevant feature, if present, receives resource and is treated the same way as the relevant feature during the decision process. In Models 2 and 5, an irrelevant feature receives resource but is ignored during the decision process. In Models 3 and 6, an irrelevant feature does not receive any resource.

The models for two-feature objects are analogous to those for one-feature objects in Chapter 2. I now describe the modifications necessary for objects with two features, say orientation and color.

2.1 From resource to encoding precision

I modeled the encoding precision of a feature of an object as the product of the amount of attentional/memory resource allocated to that object and that feature (which could be called a “top-down” factor), and a “bottom-up” factor representing low-level variables such as stimulus contrast and the width of the feature tuning curves. In line with the variable-precision model (54, 64), I allow the top-down resource factor at the i^{th} location in a stimulus array, denoted by $J'_{\text{array},i}$, to be variable across arrays (first and second), locations (objects), and trials. Specifically, I model $J'_{\text{array},i}$ as drawn independently across arrays, locations i , and trials from a gamma distribution with mean \bar{J}' and scale parameter τ ; for this stochastic process, I will use the notation

$$(\bar{J}', \tau) \rightarrow J'_{\text{array},i}$$

Then, $J'_{\text{array},i}$ is multiplied by a bottom-up factor α to yield the value of encoding precision at the i^{th} location in a given array:

$$J_{\text{array},i} = \alpha J'_{\text{array},i}.$$

I convert the drawn precision values to concentration parameters $\kappa_{\text{array},i}$ through

$$J_{\text{array},i} = \kappa_{\text{array},i} \frac{I_1(\kappa_{\text{array},i})}{I_0(\kappa_{\text{array},i})}$$

where I_1 is the modified Bessel function of the first kind of order one (61).

2.2 Step 1: Encoding

The stimuli on a given trial are represented by a quadruplet $(\boldsymbol{\theta}_{\text{ori}}, \boldsymbol{\theta}_{\text{col}}, \boldsymbol{\varphi}_{\text{ori}}, \boldsymbol{\varphi}_{\text{col}})$. When two features exist in an object, I consider three possible experimental conditions using change localization. First, both features are relevant, and a change occurs in one feature of one object (Condition B). Second, one feature is relevant and the other is irrelevant in an object (Conditions C and D). In Condition C, the irrelevant feature, denoted by a subscript “irr”, never changes, and thus, $\boldsymbol{\varphi}_{\text{irr}} = \boldsymbol{\theta}_{\text{irr}}$. In Condition D, the irrelevant feature might change. In the One-change condition, an irrelevant change of magnitude Δ_{irr} is introduced at the L^{th} location: $\varphi_{L,\text{irr}} = \theta_{L,\text{irr}} + \Delta_{\text{irr}}$. In the All-change condition, an irrelevant change of magnitude is introduced at every location: $\boldsymbol{\varphi}_{\text{irr}} = \boldsymbol{\theta}_{\text{irr}} + \Delta_{\text{irr}}$, where $\Delta_{\text{irr}} = (\Delta_{1,\text{irr}} \dots, \Delta_{N,\text{irr}})$, and $\Delta_{i,\text{irr}}$ is independently drawn from a uniform distribution for every i .

Orientation precision and color precision will be drawn from their respective distributions, described below for shared-resource models (Models 1 to 3) and independent-resource models (Models 4 to 6). The stimuli and the precisions determine the distributions from which orientation and color measurements, $(\mathbf{x}_{\text{ori}}, \mathbf{x}_{\text{col}}, \mathbf{y}_{\text{ori}}, \mathbf{y}_{\text{col}})$, are drawn. I assume that the noise corrupting the orientation and color measurements is independent across arrays, locations, and features; the assumption of independence across features is supported by a recent study (1).

2.2.1 Shared-resource models (Models 1, 2, and 3)

In the shared-resource models, orientation and color share a common pool of resource. Specifically, I assume that the sum of the mean resource of orientation and that of color is equal to a fixed value \bar{J} . Thus, I let the mean orientation resource be $\rho\bar{J}$, and

that of color $(1-\rho)\bar{J}'$, where $0<\rho<1$. I assume independent variability for orientation and color:

$$\text{Orientation: } (\rho\bar{J}', \tau'_{\text{ori}}) \rightarrow J'_{\text{ori}},$$

$$\text{Color: } ((1-\rho)\bar{J}', \tau'_{\text{col}}) \rightarrow J'_{\text{col}},$$

where τ'_{ori} and τ'_{col} are the scale parameters for orientation and color, respectively. I multiply the drawn resource values J'_{ori} and J'_{col} by bottom-up factors α_{ori} and α_{col} to obtain J_{ori} and J_{col} , respectively. This process can be simplified as

$$\text{Orientation: } (\rho\bar{J}_{\text{ori}}, \tau_{\text{ori}}) \rightarrow J_{\text{ori}},$$

$$\text{Color: } ((1-\rho)\bar{J}_{\text{col}}, \tau_{\text{col}}) \rightarrow J_{\text{col}},$$

where $\bar{J}_{\text{ori}} = \bar{J}'_{\text{ori}} \alpha_{\text{ori}}$, $\tau_{\text{ori}} = \tau'_{\text{ori}} \alpha_{\text{ori}}$, $\bar{J}_{\text{col}} = \bar{J}'_{\text{col}} \alpha_{\text{col}}$, and $\tau_{\text{col}} = \tau'_{\text{col}} \alpha_{\text{col}}$. Within the category of shared-resource models, I consider two scenarios for resource sharing: sharing across all features (Models 1 and 2), or only across relevant features (Model 3). In the former scenario, an irrelevant feature takes resource from the common pool. All shared-resource models have five parameters: \bar{J}_{ori} , \bar{J}_{col} , τ_{ori} , τ_{col} , and ρ .

2.2.2 *Independent-resource models (Models 4, 5, and 6)*

In the independent-resource models, each feature has its own resource pool. Thus, whether an irrelevant feature is encoded or not does not affect the mean precision of the relevant features. Except for the sharing of resource, the independent-resource models are identical to the shared-resource models. Thus, the precision of orientation and color follow:

$$\text{Orientation: } (\bar{J}_{\text{oi}}, \tau_{\text{oi}}) \rightarrow J_{\text{oi}},$$

$$\text{Color: } (\bar{J}_{\text{cd}}, \tau_{\text{cd}}) \rightarrow J_{\text{cd}}.$$

In Models 4 and 5, irrelevant features have a pool of resource, whereas in Model 6, they do not. All independent-resource models have four parameters: \bar{J}_{ori} , \bar{J}_{col} , τ_{ori} , and τ_{col} .

2.3 Step 2: Inference

2.3.1 Two relevant features (Condition B)

In Condition B, each object has orientation and color, which are both relevant. On each trial, the change happens either in orientation or in color, with equal probabilities. Thus, the observer has to consider the hypothesis that orientation has changed and color has not, and the alternative hypothesis that color has changed and orientation has not. These hypotheses have the same a priori probability (0.5). Thus, Eq. (2.12) for the decision variable changes to:

$$d_L = \frac{1}{2} \left(\frac{p(x_{L,\text{ori}}, y_{L,\text{ori}} | C_{L,\text{ori}} = 1) p(x_{L,\text{col}}, y_{L,\text{col}} | C_{L,\text{col}} = 0) + p(x_{L,\text{ori}}, y_{L,\text{ori}} | C_{L,\text{ori}} = 0) p(x_{L,\text{col}}, y_{L,\text{col}} | C_{L,\text{col}} = 1)}{p(x_{L,\text{ori}}, y_{L,\text{ori}} | C_{L,\text{ori}} = 0) p(x_{L,\text{col}}, y_{L,\text{col}} | C_{L,\text{col}} = 0)} \right)$$

This can be simplified to

$$\begin{aligned} d_L &= \frac{1}{2} \left(\frac{p(x_{L,\text{ori}}, y_{L,\text{ori}} | C_{L,\text{ori}} = 1)}{p(x_{L,\text{ori}}, y_{L,\text{ori}} | C_{L,\text{ori}} = 0)} + \frac{p(x_{L,\text{col}}, y_{L,\text{col}} | C_{L,\text{col}} = 1)}{p(x_{L,\text{col}}, y_{L,\text{col}} | C_{L,\text{col}} = 0)} \right) \\ &= \frac{1}{2} (d_{L,\text{ori}} + d_{L,\text{col}}), \end{aligned}$$

where

$$d_{i,\text{ori}} = \frac{I_0(\kappa_{i,\text{ori},x}) I_0(\kappa_{i,\text{ori},y})}{I_0 \left(\sqrt{\kappa_{i,\text{ori},x}^2 + \kappa_{i,\text{ori},y}^2 + 2\kappa_{i,\text{ori},x}\kappa_{i,\text{ori},y} \cos(x_{i,\text{ori}} - y_{i,\text{ori}})} \right)}$$

and analogous for color. Thus, under the generative model used here, the decision variable is simply the average of the decision variables for the individual features. The reported location is

$$\hat{L} = \underset{L}{\operatorname{argmax}}(d_{L,\text{ori}} + d_{L,\text{col}}).$$

2.3.2 *One relevant and one irrelevant feature (Conditions C and D)*

If an irrelevant feature is encoded and taken into account during decision-making (Conditions C and D), I assume that the decision variable is similar to that of Condition B:

$$d_L = \frac{1}{2}(d_{L,\text{rel}} + d_{L,\text{irr}}),$$

where $d_{L,\text{rel}}$ is the decision variable of the relevant feature, and $d_{L,\text{irr}}$ is that of the irrelevant feature. If an irrelevant feature is encoded but ignored during decision-making, for Conditions C and D, the decision variable is similar to that of Condition A (single feature): $d_L = d_{L,\text{rel}}$.

2.4 Step 3: Response probabilities

For each model for two-feature objects, I calculated the response probabilities as follows. Each two-feature trial is characterized by change magnitude vectors of orientation and color, denoted by Δ_{ori} and Δ_{col} , respectively, which depend on the condition. (For example, in Condition C, one of the vectors is the zero vector, whereas in Condition D, neither is.) In each experimental condition, I estimated the model observer's probability of a correct response, $p(\text{correct} | \Delta_{\text{ori}}, \Delta_{\text{col}}, \omega)$, through Monte Carlo simulations with 1280 samples of $\kappa_{\text{ori},x}$, $\kappa_{\text{ori},y}$, \mathbf{x}_{ori} , \mathbf{y}_{ori} , $\kappa_{\text{col},x}$, $\kappa_{\text{col},y}$, \mathbf{x}_{col} , and \mathbf{y}_{col} . This was done in analogy to 2.1.4 of the current chapter.

In conclusion, combining the possible answers to Q1 and Q2 results in six models (Fig. 4.1). In Models 1 to 3, orientation and color share memory resource, whereas in Models 4 to 6, orientation and color have independent pools of resource. In

Models 1 and 4, an irrelevant feature, if present, receives resource and is treated the same way as the relevant feature during the decision process. In Models 2 and 5, an irrelevant feature receives resource but is ignored during the decision process. In Models 3 and 6, an irrelevant feature does not receive any resource. After ruling out of these all models but one, I will consider Q3 separately.

Model	Memory resource between features	Irrelevant Feature		Condition A		Condition B			Condition C			Condition D			Model ruled out by
				one feature	two relevant features	a relevant feature + a non-changing irrelevant feature	a relevant feature + a changing irrelevant feature	RelF resource	Perf.	RelF ₁ resource	RelF ₂ resource	Perf.	RelF resource	Effect of IrrF on decision	
1	Shared	Encoded	Affects decision		ref.					Small			Large		Exp. 2
2	Shared	Encoded	Ignored		ref.					None			None		Exp. 2
3	Shared	Not encoded	N/A		ref.					None	Same		None	Same	Exp. 1
4	Independent	Encoded	Affects Decision		ref.			Same		Small			Large		Exp. 3
5	Independent	Encoded	Ignored		ref.			Same		None	Same		None	Same	
6	Independent	Not encoded	N/A		ref.			Same		None	Same		None	Same	Exp. 4

Exp. 1 Exp. 1
 Exp. 2 Exp. 2
 Exp. 3 Exp. 3

Exp. 4
 Exp. 2

A common pool of memory resource (portion: or) RelF: relevant feature
 Resource pool of the first relevant feature IrrF: irrelevant feature
 Resource pool of the second relevant feature Perf.: prediction for subject performance relative to that in Condition A (ref.)
 Performance lowered by resource-sharing
 Performance lowered by changes in irrelevant feature (small effect: large effect:)

Figure 4.1 Models of VSTM for Q1 and Q2 and their predictions for performance in different conditions.

3 Methods

3.1 Laboratory experiments

3.1.1 Stimuli

An orientation-only stimulus was a white ellipse of luminance of 95.7 cd/m^2 with minor and major axes of 0.41 and 0.94 degrees of visual angle (deg), respectively. A color-only stimulus was a disc with a diameter of 0.62 deg, with color drawn from 360 values uniformly distributed along a circle in the fixed- L^* plane of CIE 1976 (L^* , a^* , b^*) color space corresponding to a luminance of 95.7 cd/m^2 , with center $(a^*, b^*)=(12, 13)$ and radius 60. A two-feature stimulus was a colored ellipse.

Stimuli were displayed on a 19" LCD monitor at a viewing distance of approximately 60 cm. Stimuli were presented on a mid-level grey background of luminance 33.1 cd/m^2 . Stimuli were equally spaced along an imaginary circle of radius 7 degrees of visual angle (deg) around fixation (calculated assuming a viewing distance of exactly 60 cm), at angles $i/N \cdot 360^\circ$, where $i = 1, \dots, N$, and $N = 4$ or 8. In the Separate condition of Experiment 5, color stimuli were positioned at even i , orientation stimuli at odd i . All experiments were programmed using Psychophysics Toolbox in MATLAB (76, 77).

3.1.2 Subjects

Experiment 1: eight subjects (one author), age range 23–30. Experiment 2: eight subjects (one author), age range 23–30. Experiment 3: five subjects (one author), age range 23–30. Experiment 5: five subjects (one author), age range 26–30. Besides the author, all subjects were naïve to the goals of the experiments. All subjects gave

informed consent. The experimental protocol was approved by the IRB of Baylor College of Medicine.

3.1.3 Conditions

3.1.3.1 Condition A: One feature

The trial sequence consisted of the presentation of a fixation cross (1000 ms), the first stimulus array (100 ms), a delay period in which the fixation cross was present (1000 ms), the second stimulus array (100 ms), and a response screen (present until response). The second array was identical to the first except that exactly one object was different in its feature value from the first array. The magnitude of the change was drawn from a uniform distribution. Each object had an equal probability of changing. The response screen consisted of empty circles at the same locations as where objects were presented in the stimulus arrays. The task was to click on the location of the object that had changed. After the response, feedback was provided: the fixation cross turned green if the response was correct, red if the response was incorrect. At the beginning of each session, I instructed subjects that their task was to localize the change and that the change could be small or large.

3.1.3.2 Condition B: Two relevant features

Condition B differed from Condition A in that all objects had two features (see Methods). The change occurred either in orientation or color with equal probability. At the beginning of each session, I instructed subjects that their task was to localize the change in either feature and that the change could be small or large.

3.1.3.3 Condition C: One relevant feature and one non-changing irrelevant feature
Condition C was identical to Condition A except that all objects also had an irrelevant feature. The irrelevant feature did not change between the two arrays. At the beginning of each session, I instructed subjects that their task was to localize the change in the relevant feature, and that the change could be small or large.

3.1.3.4 Condition D: One relevant feature and one changing irrelevant feature
Condition D differed from Condition B in following ways. In the “One-change” experiment, on each trial, one object (chosen with equal probabilities) changed orientation and another object (independently chosen, also with equal probabilities) changed color. Both changes were independently drawn from uniform distributions. By chance, both changes could occur in the same object. One feature was relevant and the other was irrelevant. Which feature was relevant and which feature was irrelevant remained fixed throughout a session. The task was to click on the location of the object that had changed in its relevant feature. At the beginning of each session, I instructed subjects that their task was to localize the change in the relevant feature, that the change could be small or large, that one randomly chosen object would change in its irrelevant feature, and that by chance the same object could change in both features.

The “All-change” experiment was identical to the One-change experiment except that on each trial in the second array, *every* object (not just one) changed in its irrelevant feature. All changes in irrelevant features as well as the change in the relevant feature were independently drawn from a uniform distribution. At the beginning of each session, I instructed subjects that their task was to localize the change in the relevant feature, that the change could be small or large, and that all objects would change in their irrelevant features.

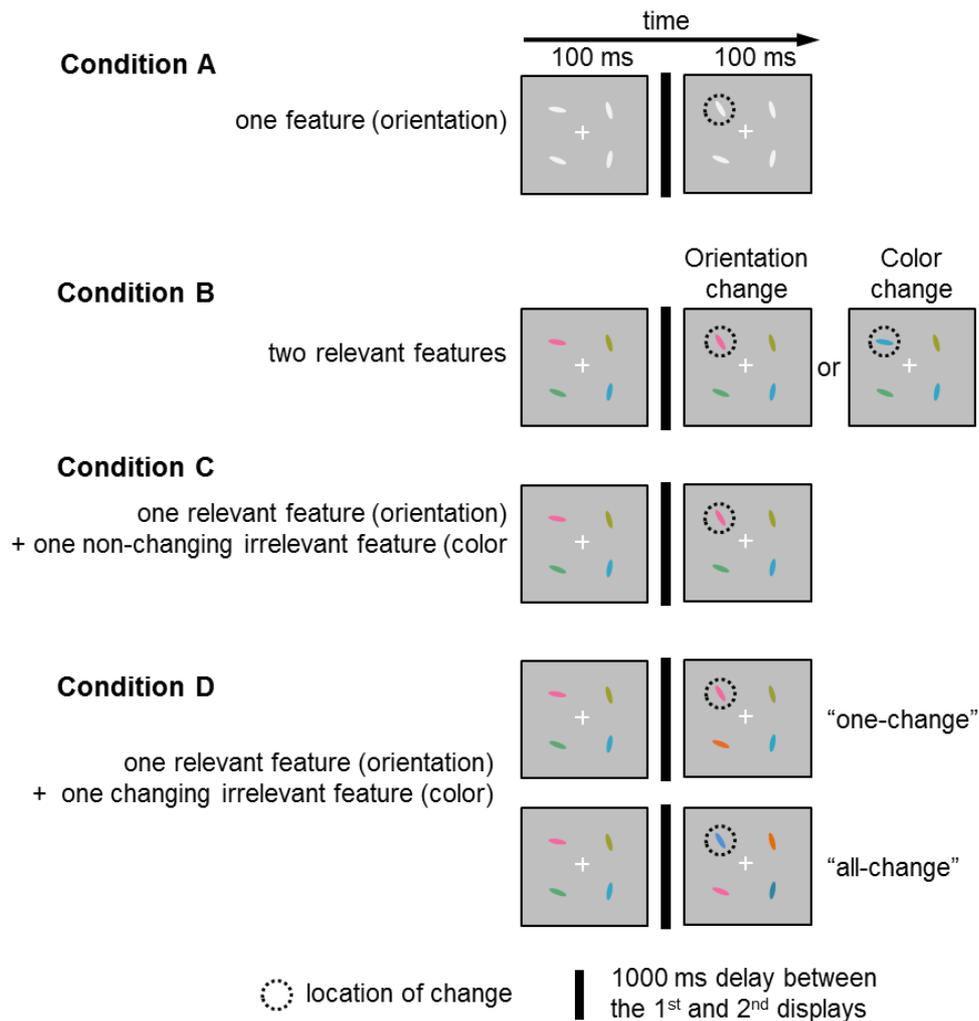


Figure 4.2 Trial procedures in Experiments 1, 2, and 3. Subjects click on the location where a (relevant) change has occurred. (A) Orientation trial of Condition A: orientation is the only feature. (B) Condition B: stimuli have two features and the change has occurred in either. (C) Orientation trial of Condition C: stimuli have two features but only orientation is relevant. The irrelevant feature (here color) does not change. (D) Orientation trial of Condition D: as C, but the irrelevant feature value (here color) has changed in either one or all stimuli.

3.1.4 Experiment 1: Conditions B and C

Experiment 1 consisted of four sessions, run on different days: two sessions of Condition B (statistically identical to each other) and two sessions of Condition C (of which one was a relevant-orientation session and one a relevant-color session). The order of the sessions was random for each subject. Each session consisted of four blocks of 150 trials. Hence, each subject completed $4 \cdot 4 \cdot 150 = 2400$ trials in total. At the beginning of each session, subjects completed 10 practice trials. Each session lasted about 45 min. To test consistency across different set sizes, I conducted the same experiment at a higher set size ($N=8$), using 8 colored oriented ellipses.

3.1.5 Experiment 2: Conditions A and C

Experiment 2 consisted of four sessions, run on different days: two sessions of Condition A (a relevant-orientation session and a relevant-color session) and two sessions of Condition C (a relevant-orientation session and a relevant-color session). Set size was 4. Otherwise, Experiment 2 was the same as Experiment 1.

3.1.6 Experiment 3: Conditions C and D

Experiment 3 consisted of six sessions: two sessions of Condition C (a relevant-orientation session and a relevant-color session), two sessions of One-change Condition D (same), and two sessions of All-change Condition D (same). Otherwise, Experiment 3 was the same as Experiment 1.

3.2 Amazon Mechanical Turk experiments

3.2.1 Stimuli

Set size was 1 (single colored oriented ellipse). The stimulus was a colored ellipse presented in the center of a rectangular 400×400-pixel mid-level grey window. Minor and major axes were 16 and 7 pixels, respectively. Possible colors were as in the laboratory experiments, except that I could not control luminance. The experiments were programmed using HTML, CSS, and Raphaël library in JavaScript.

3.2.2 Subjects

The two experiments (orientation and color) were conducted online on www.mturk.com. Subjects enrolled by selecting our experiment from a list of “Human Intelligence Tasks”. To maximize consistency in response modality and viewing conditions, I prevented enrollment via mobile devices such as phones or tablets. To ensure normal color vision, I asked subjects to take a 6-item Ishihara Color Test. If subjects failed the test, subjects could not continue and were blocked from later enrollment in either experiment. Taking this selection into account, 600 subjects participated in the color experiment, and 600 in the orientation experiment. I prevented overlap between these groups by blocking user IDs. Each subject was paid 25 cents. The experimental protocol was approved by the IRB of Baylor College of Medicine.

3.2.3 Procedure for Experiment 4 (delayed estimation)

Experiment 4 consisted of two experiments (orientation and color), which were both conducted on Amazon Mechanical Turk (www.mturk.com) without direct human supervision. Once a subject selected the experiment, they were told to enlarge the window size of their web browser to at least 800×600 pixels. If they did not do this,

they could not continue. Then, the subject completed a six-item Ishihara Color Test to test for color vision deficiencies. A subject who passed this test was then led step by step through an example trial accompanied by on-screen instructions. This instruction phase was self-paced and subjects could freely move back and forth between the screens of the example trial.

The trial sequence consisted of the presentation of a fixation cross (1000 ms), the stimulus (150 ms), a delay period with a blank screen (1000 ms), and a response screen, which lasted until the subject responded. The stimulus screen consisted of a single colored ellipse, presented centrally. Each experiment lasted approximately 5 minutes.

The first experiment consisted of 30 relevant-orientation trials, followed by a surprise relevant-color trial. On the first 30 trials, the response stage consisted of a response probe appearing with a message asking subjects to report the orientation. The probe was a colored ellipse with the same color as the stimulus and an orientation drawn from a uniform distribution. Subjects could rotate the ellipse by moving a mouse, and submitted their response by pressing the spacebar. Feedback was then provided by presenting the original stimulus (“correct”) and the reported stimulus (“reported”) simultaneously in the top and bottom parts of the stimulus window, respectively. On the 31st (surprise) trial, after the delay period, a color wheel appeared with a message asking subjects to report the color of the stimulus by clicking on the color wheel. No feedback was given on this trial, and the experiment terminated immediately afterwards. The second experiment was identical to the first experiment except that the first 30 trials were relevant-color trials and the surprise trial was a relevant-orientation trial.

4 Results

For the ANOVA used to test for the effect of change magnitude on accuracy, we had to bin change magnitude. This raises the question of whether bin size matters. We tested six different bin sizes (5, 6, 7.5, 10, 15, 30°), and found a highly significant effect of change magnitude on accuracy ($p < 0.001$) for each. Here and elsewhere, I binned change magnitude into 10° bins.

4.1 Answering Q1: Resource is not shared between orientation and color.

Psychometric curves from [Experiment 1](#) are shown in Fig. 4.3. For orientation, a two-way repeated-measures ANOVA reveals a significant main effect of change magnitude ($F(8, 56) = 81.98, p < 0.001$), no significant main effect of condition ($F(1, 7) = 0.16, p = 0.70$), and no significant interaction ($F(8, 56) = 0.34, p = 0.95$). This is consistent for color (respectively: $F(8, 56) = 104.46, p < 0.001$; $F(1, 7) = 1.88, p = 0.21$; $F(8, 56) = 1.00, p = 0.44$). Model 3 postulates that resource is shared between orientation and color, and that an irrelevant feature, if present, is not encoded. It predicts a difference between Conditions B and C. Thus, I find no evidence for Model 3.

Since a frequentist test cannot prove a null result, I also performed formal model comparison. I compared Model 3 against Model 5/6 (Models 5 and 6 make the same prediction for this experiment). To implement these models, I need to choose a specific single-feature encoding model. The item-limit model (28, 38), on which the formulation of hypotheses of H1 and H2 has traditionally been based, cannot be used, since it does not incorporate memory noise and therefore does not predict any dependence of performance on change magnitude. Therefore, I instead use the variable-precision encoding model (64), which is the simplest well-fitting single-feature encoding model (54). Given the qualitative pattern of results, I do not expect our conclusion to depend

on this choice, and I refrain from making any claims about single-feature encoding models in this paper. I found that for every individual subject, the Akaike Information Criterion (AIC) of Model 3 is higher than that of Models 5/6, on average by 161 ± 32 (mean \pm s.e.m.). This is consistent with an identical experiment with set size 8 (AIC difference: 81 ± 19 , see SI Text). Therefore, I can reject Model 3.

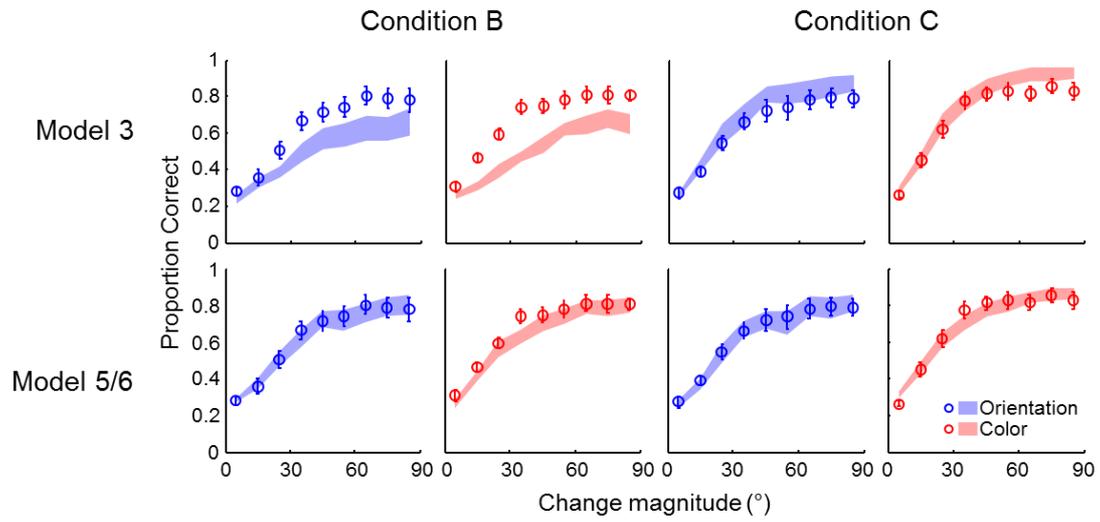


Figure 4.3 Experiment 1. Proportion correct as a function of change magnitude (8 subjects), with model fits (shaded areas: mean \pm s.e.m.). Performance was indistinguishable between Conditions B and C, in accordance with Model 5/6 but not with Model 3.

I cannot rule out yet that resource is shared, since it is possible that the irrelevant feature receives a portion of the shared resource, as in Models 1 and 2. To rule these models out, I conducted Experiment 2, in which subjects were tested on Condition C (a relevant and a non-changing irrelevant feature) as well as a new condition (Condition A in Fig. 4.2), in which stimuli had only one feature. In blocks of orientation trials, the stimuli were four gray ellipses, one of which had changed its orientation. In blocks of color trials, stimuli were four colored discs, one of which had changed its color. Models 1 and 2 both postulate that memory resource is shared between the features in Condition C, and therefore predict lower performance in Condition C than in Condition A. (Moreover, in Model 1, the irrelevant feature is taken into account in the decision process, thereby adding noise and further reducing performance.)

Psychometric curves are shown in Fig. 4.4. For orientation, a two-way repeated-measures ANOVA reveals a significant main effect of change magnitude ($F(8, 56) = 136.17, p < 0.001$), no significant main effect of condition ($F(1, 7) = 0.04, p = 0.84$), and no significant interaction ($F(8, 56) = 1.60, p = 0.15$). This is consistent for color (respectively: $F(8, 56) = 111.77, p < 0.001$; $F(1, 7) = 0.002, p = 0.97$; $F(8, 56) = 1.16, p = 0.34$). Thus, we find no evidence for Models 1 or 2.

For every individual subject, the AICs of Models 1 and 2 are higher than that of Models 5/6, on average by 210 ± 13 and 154 ± 11 , respectively (Table 4.1). Thus, I can reject Models 1 and 2.

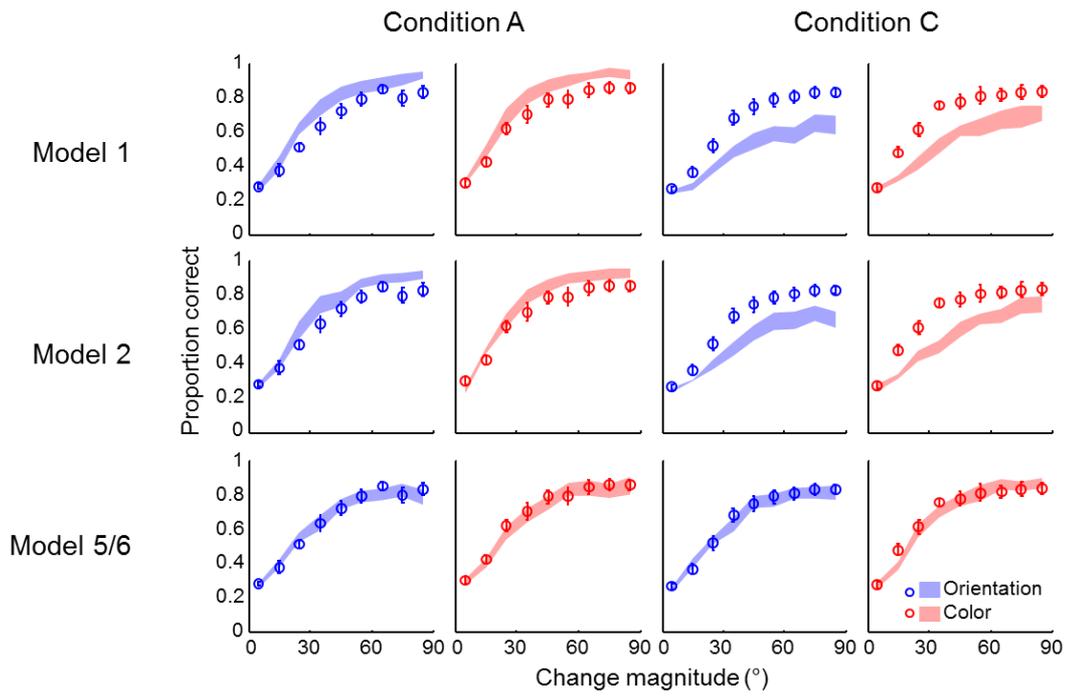


Figure 4.4 Experiment 2. Proportion correct as a function of change magnitude (8 subjects), with model fits (shaded areas: mean \pm s.e.m.). Performance was indistinguishable between Conditions A and C, in accordance with Model 5/6 but not with Models 1 and 2.

4.2 Answering Q2: The irrelevant feature does not affect performance.

To test whether the irrelevant feature affects performance, I followed the concept of Hyun et al. (70) and introduced changes in the irrelevant feature in Experiment 3. I tested subjects on Condition C (a relevant and a non-changing irrelevant feature) and a new condition (Condition D), in which the value of the irrelevant feature of an object had changed. In the “One-change” version of this condition, one object had changed in its irrelevant feature. In the “All-change” version, every object had changed in its irrelevant feature.

Among the models not yet ruled out, Models 5 and 6 both predict that performance in both conditions is identical, while Model 4 predicts that any change in the irrelevant feature will decrease performance, and the more objects change in their irrelevant feature, the greater the decrease. Therefore, Model 4 predicts that performance is lower in the One-change version of Condition D than in Condition C, and even lower in the All-change version of Condition D.

Psychometric curves are shown in Fig. 4.5. For orientation, a two-way repeated-measures ANOVA reveals a significant main effect of change magnitude ($F(8, 32) = 179.69, p < 0.001$), no significant main effect of condition ($F(2, 8) = 0.94, p = 0.43$), and no significant interaction ($F(16, 64) = 1.68, p = 0.07$). This is consistent for color (respectively: $F(8, 32) = 240.20, p < 0.001$; $F(2, 8) = 0.11, p = 0.90$; $F(16, 64) = 1.51, p = 0.13$). Thus, I find no evidence for Model 4.

When I assume that in Model 4, a feature is encoded with the same precision regardless of whether it is relevant or irrelevant, then for every individual subject, the AIC of Model 4 is higher than of Model 5/6 for every subject, on average by $(6.8 \pm 1.5) \cdot 10^2$ (Table 4.1). Therefore, I can reject that, for our stimuli, the irrelevant

feature is both encoded and treated like the relevant feature during decision-making. The absence of an effect of an (either changing or non-changing) irrelevant feature on performance is consistent with classic change detection studies that did not vary change magnitude (31, 50, 68, 69).

Mean and s.e.m. of maximum-likelihood estimates of the models from Experiments 1, 2, and 3 are shown in Tab. 3.2.

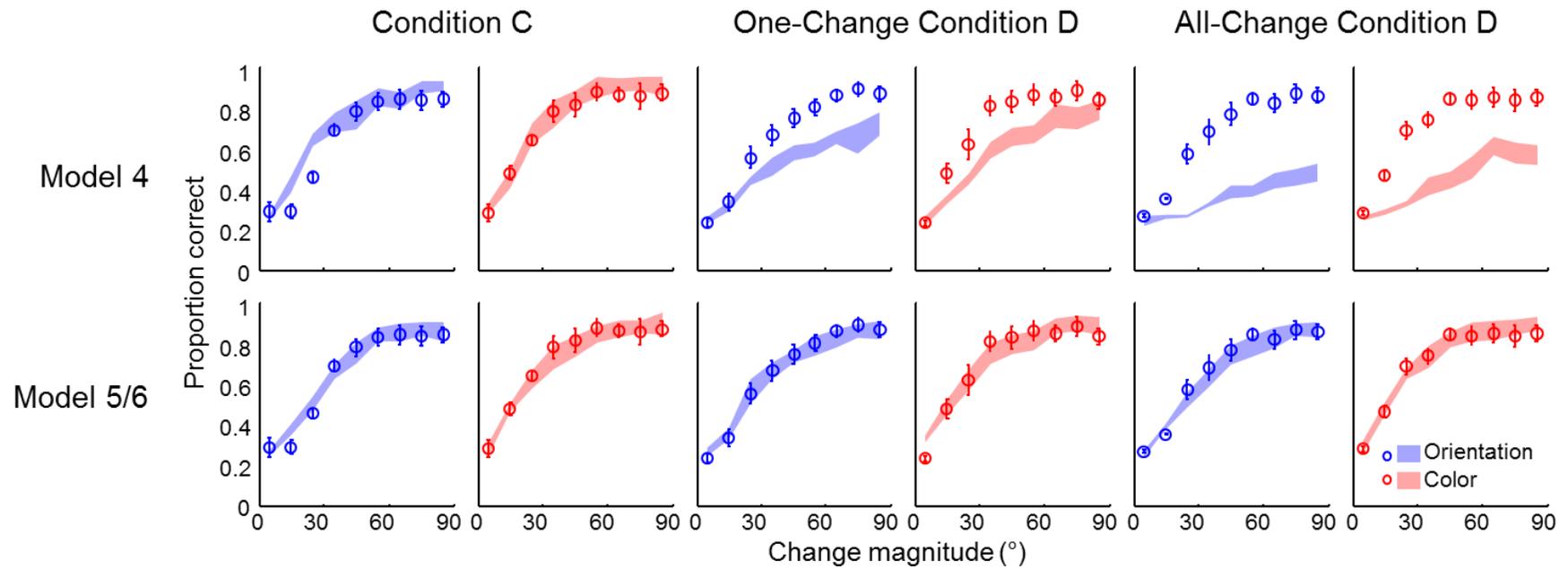


Figure 4.5 Proportion correct as a function of change magnitude (5 subjects), with model fits (shaded areas: mean \pm s.e.m.). Performance was indistinguishable between Conditions C and D, in accordance with Model 5/6 but not with Model 4.

Experiment	Model	Parameters				
		\bar{J}_{ori}	\bar{J}_{col}	τ_{ori}	τ_{col}	ρ
1 ($N=4$)	3	73.9±9.9	89±12	32.0±6.9	37±10	0.501±0.026
	6	92±21	184±36	63±15	133±31	
1 ($N=8$)	3	78.5±5.3	93.1±6.1	34.5±3.8	38.5±7.6	0.441±0.035
	6	79.2±10.7	151±30	45.6±7.5	89±19	
2	1	72.4±8.4	110±18	30.6±8.0	41.9±7.3	0.500±0.020
	2	65±10	92.6±7.6	22.4±3.6	38±11	0.513±0.025
	5	79±11	110±25	51.3±8.2	80±28	
3	4	70.3±7.7	113±13	31.1±6.2	51.6±9.8	
	5	45.7±5.1	138±10	20.1±5.5	87±21	

Table 4.1 Mean and s.e.m. of maximum-likelihood estimates of parameters.

Experiment	Model	Subject index								Mean
		1	2	3	4	5	6	7	8	
1 ($N=4$)	3	2253	2895	3364	3039	2723	3193	2548	3257	2909
	6	1990	2796	3305	2856	2641	2951	2274	3165	2747
1 ($N=8$)	3	3148	2865	3167	2994	2785	2911	2849	3022	2967
	6	2996	2814	3001	2887	2732	2856	2827	2975	2886
2	1	3009	3051	2431	3180	2674	3068	2241	2560	2777
	2	2954	3004	2379	3133	2616	3025	2172	2484	2721
	5	2773	2830	2203	3049	2459	2878	2019	2322	2567
3	4	4391	4277	4247	3976	4658				4310
	5	3787	3626	3646	2738	4355				3630

Table 4.2 Individual AIC values of all models from Experiments 1, 2, and 3.

4.3 Answering Q2's follow-up question: The irrelevant feature is encoded but ignored.

The remaining models, Models 5 and 6, make different claims: “the irrelevant feature is encoded but ignored during decision-making” (Model 5) and “the irrelevant feature is not encoded” (Model 6). These models cannot be distinguished unless I probe the encoding more directly; this same problem applies to classic change detection studies that claimed that the irrelevant feature is not encoded (31, 50).

I address this problem as directly as possible in Experiment 4, using a delayed-estimation task (39). On each trial, the stimulus was a colored ellipse, with orientation and color independently drawn from their respective uniform distributions (Fig. 4.6). After a delay, subjects were asked to recall one of the feature values. The subject's response is taken to be a proxy of their memory of the stimulus, and thus the task has only a minimal decision-making component. In the first thirty trials, the task was to recall the value of the relevant feature; across these trials, which feature was relevant was kept fixed. On the surprise 31st trial, an on-screen message instructed subjects to recall the value of the *irrelevant* feature of the stimulus that they just saw. Model 6 predicts that subjects will be guessing on the surprise trial and that their estimation errors will be distributed uniformly.

Once a subject has experienced a surprise trial, the irrelevant feature suddenly becomes relevant. Therefore, each subject can be tested on only one irrelevant-feature trial. To solve this problem, I conducted the experiment on Amazon Mechanical Turk (Mturk), an online platform for data collection. I crowd-sourced data from a large group of online subjects.

For both orientation and color, the pooled distribution of the estimation error on the relevant-feature trials is significantly different from uniform (Kolmogorov-Smirnov test: orientation: $D = 0.35$, $p < 0.001$, Fig. 4.7A; color: $D = 0.36$, $p < 0.001$, Fig. 4.7B). Critically, the error distribution on irrelevant-feature trials is not uniform either (color: $D = 0.28$, $p < 0.001$, Fig. 4.7A; orientation: $D = 0.09$, $p < 0.001$, Fig. 4.7B). This rejects Model 6 and suggests that subjects do encode the irrelevant feature.

The circular standard deviation (78) was 13.7° for relevant-orientation trials, 55.1° for irrelevant-orientation trials, 23.3° for relevant-color trials, and 51.6° for irrelevant-color trials. This shows that subjects' responses were less precise when they recalled the irrelevant feature. (Circular means were small: -0.3° , -2.3° , -0.5° , and -3.3° , respectively.)

Experiment 4 narrows the possible models down to Model 5 only: orientation and color have independent pools of resource, and irrelevant features are encoded but ignored during decision-making.

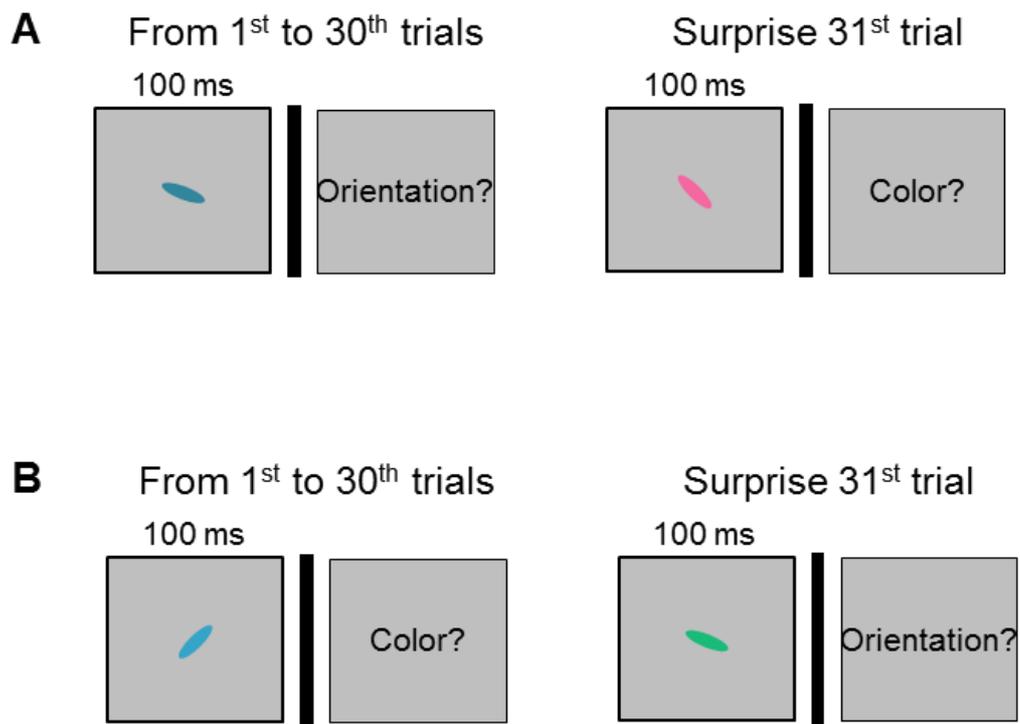


Figure 4.6 Delayed estimation task. Subjects probe an estimate of the relevant feature for the first 30 trials, and on the surprise 31st trial, they are asked to probe the irrelevant feature.

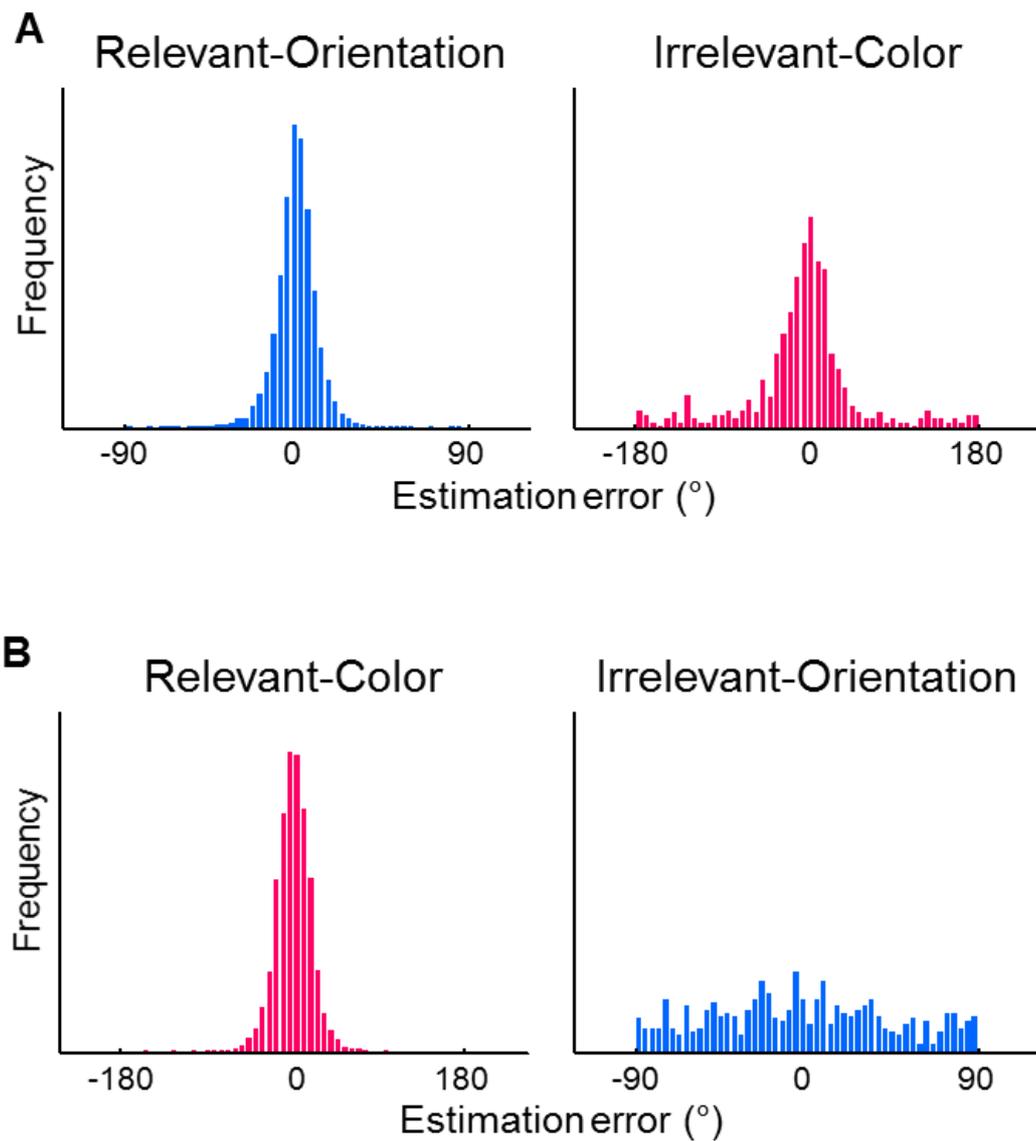


Figure 4.7 Experiment 4. The error distribution from pooled data (600 subjects respectively) is shown below the procedure. **(A)** Relevant feature: orientation, irrelevant feature: color. Both distributions are significantly different from uniform. **(B)** Relevant feature: color, irrelevant feature: orientation. Both distributions are significantly different from uniform.

5 Discussion

Most objects in nature consists of multiple features. To understand how our brain processes this information using VSTM, in this chapter, I conducted a series of psychophysics experiments using objects with orientation and color, and applied the continuous-resource view from previous chapters.

The traditional object-based vs. feature-based approach on VSTM for objects with multiple features is based on the discrete-capacity view, which was proven to be less accurate describing human VSTM behavior than the continuous-resource models in Chapters 2 and 3. In this chapter, I reinterpreted the questions from earlier studies on the unit of VSTM (H1 and H2), which are based on the noise view, and established models based on the following issues: 1) Is memory resource shared by orientation and color? 2) Is an irrelevant feature encoded? 3) If yes, is the encoded irrelevant feature ignored during decision-making? Considering the combinations of the answers to the questions, I came up with six models. Experiments 1-4 systematically rule out all suggested models but one (Model 5).

Experiments 1 and 2 answered the first question. The results show that orientation and color does not share a VSTM resource. Experiment 3 shows that irrelevant feature does not affect performance, but I found that subjects can remember irrelevant feature in Experiment 4, which concludes that an irrelevant feature is encoded but it does not affect performance since it can be ignored during decision-making.

The results of Experiments 1 and 2 are consistent with the classic change detection studies that did not vary change magnitude (28, 31, 33, 66) and our conclusion parallels their conclusion that each feature has a separate capacity (33, 66), except that we replace “item capacity” by “memory resource”. These result provides a counterpoint

to the mixed results obtained in another study that worked within the noise-based framework (41): in a delayed-estimation task, color memories were somewhat noisier when both orientation and color had to be remembered than when only color had to be remembered. However, orientation memory did not suffer such a cost, but it did when orientation stimuli were not chosen independently of each other within a display. A change detection experiment aimed at addressing the same question was equally inconclusive: a cost for memorizing two features over one was observed when the change was small but not when it was large (41). By contrast, the results are consistent across features and change magnitudes.

The results in this chapter suggest that orientation and color do not share VSTM resource, and an irrelevant feature is encoded but ignored during decision-making (Model 5).

Chapter 5

VSTM for two-feature objects (2): spatial allocation of memory resource

1 Introduction

In Chapter 4, I discussed the allocation of a feature resource (orientation or color) to objects by using stimuli where the spatial allocation of observers' attention was constant across conditions. In addition to the questions on object-based VSTM that I introduced there (H1: object-based VSTM is not affected by addition of features in the same object, and H2: irrelevant feature is encoded with relevant feature), another question regarding object-based VSTM has been asked in literature has a closer relationship to spatial attention.

H3. If VSTM is object-based, then memorizing two features of the same object should be easier than of two different objects.

Such an "object benefit" (37) can be tested for by comparing the following two conditions: 1) observers memorize N objects, and each object has two features; 2) the number of objects observers memorize is now doubled ($2N$), but each object now has only one feature. In both cases, the total number of the features in a display is the same as $2N$. Earlier studies found that performance was found to be lower in the latter condition (32, 66, 72). Based on the results, Olson and Jiang (66) suggested that VSTM is "weakly" object-based. However, the fact that spatial attention is divided over more objects in the second condition complicates this conclusion.

Earlier studies on this question had similar problems in studies on H1 and H2: they used highly discriminable colors and avoided the notion of noise in memory, and there was no discussion on decision-making step of change detection. Similar to Chapter 4, therefore, I apply the continuous-resource view of memory to H3, and translate it to a new question:

(Q3) Spatial allocation of resource. Is the resource available for a given feature allocated only to the objects that have that feature?

In this chapter, I answer Q3. I examine how a feature resource is allocated to locations by varying the number of objects in stimuli. I conducted change localization experiments in conditions where the number of objects or features varies systematically. I found that orientation and color resources are distributed to locations in a partially dependent manner.

2 Theory and models

I suggest two models: the No-leak and Leak models. For both models, I use the assumptions about encoding precision and decision-making of Model 5 from Chapter 4: each feature has its own resource, and irrelevant features are encoded but ignored in decision-making. In the No-leak model, the entire amount of each feature's resource is distributed to relevant locations. In the Leak model, only a portion of the resource is distributed to relevant locations.

Each model makes predictions for three experimental conditions: Together- N (identical to Condition B), Separate, and Together- $2N$. The first two conditions have N objects, and the last has $2N$ objects. To fit the data from all conditions simultaneously, we model mean precision as $\frac{\bar{J}_{N=1}}{N}$, where N is set size and $\bar{J}_{N=1}$ is mean precision at set size 1. The details of the Separate and Together- $2N$ conditions are described in section 3.1 of the current chapter.

2.1 No-leak model

The No-leak model predicts the same behavior in the Together- N condition as in Condition B. For the Separate condition, per feature, the entire amount of memory

resource is assigned only to the relevant locations ($N=4$). We have 8 decision variables: four from orientation and four from color. This is comparable to a change localization task with 8 one-feature objects, although here there are four color objects and four orientation objects. Finally, the model's predictions for the Together- $2N$ condition are identical to those for the Together- N , except that encoding precision is lower because $N=8$. The No-leak model has four parameters: \bar{J}_{ori} , \bar{J}_{col} , τ_{ori} , τ_{col} .

2.2 Leak model

The leak model is identical to the No-leak model except that in the Separate condition, per feature, some amount of resource “leaks away” from relevant objects, which reduces the mean precision for that feature in those objects to $(1-r)\frac{\bar{J}_{N=1}}{N}$, where r is a feature-specific leak parameter ($0 < r < 1$), and $N=4$. The model predictions for the Together- N and Together- $2N$ conditions are identical to those of the No-leak model. The leak model has six parameters: \bar{J}_{ori} , \bar{J}_{col} , τ_{ori} , τ_{col} , r_{ori} , and r_{col} .

3 Methods

3.1 Psychophysics

Stimuli are identical to those from Chapter 4. Five subjects (one author) participated and, the age ranges from 26 to 30. Besides the author, all subjects were naïve to the goals of the experiments. We examined Q3 using three conditions (Fig. 5.1) that paralleled those in (66), namely “Together- N ”, “Separate”, and “Together- $2N$.”

The experiment consisted of two sessions of each condition, in total of six. Each session was run on different days. The order of the sessions was random for each subject.

Each session consisted of four blocks of 150 trials. Hence, each subject completed $6 \cdot 4 \cdot 150 = 3600$ trials in total.

3.2 Experimental conditions

3.2.1 *The Together-N: four two-feature objects*

In the “Together- N ” condition, each of the four objects had both orientation and color, and the change had occurred with equal probabilities in either feature (this is the same as Condition B in Chapter 4).

3.2.2 *The Separate: four orientation objects and four color objects*

In the “Separate” condition, the stimuli were four discs with colors independently drawn from a uniform distribution and four gray ellipses with orientations independently drawn from a uniform distribution, for a total of 8 objects. The locations of discs and ellipses remained the same through across trials.

3.2.3 *The Together-2N: eight two-feature objects*

The “Together- $2N$ ” condition was identical to the Together- N condition except that set size was 8.

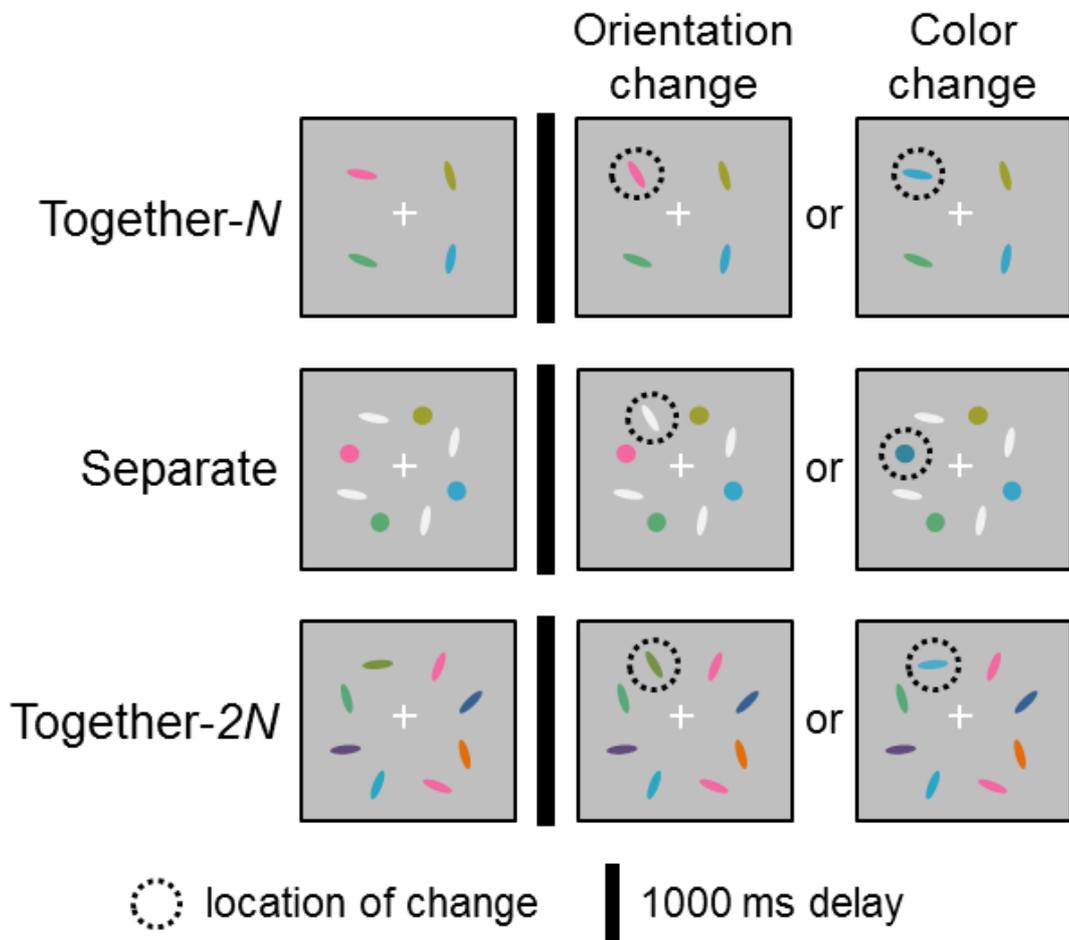


Figure 5.1 Trial procedures. A change occurred either in orientation or color. Subjects click on the location where the change occurred.

4 Results

4.1 Summary statistics

Psychometric curves are shown in Fig. 5.2. Performance was higher in the Together- N condition than in the Separate condition, and higher in the Separate condition than in the Together- $2N$ condition. For orientation, a two-way repeated-measures ANOVA shows a significant main effect of change magnitude ($F(8, 32) = 62.14, p < 0.001$), a significant main effect of condition ($F(2, 8) = 82.92, p < 0.001$), and significant interaction ($F(16, 64) = 5.02, p < 0.001$). This is consistent for color (respectively: $F(8, 32) = 70.13, p < 0.001$; $F(2, 8) = 82.92, p < 0.001$; $F(16, 64) = 2.13, p = 0.02$). However, this qualitative pattern of results can be predicted both by the Leak and the No-leak models. This is easiest to understand from the fact that in the Separate condition, there are eight instead of four locations, and therefore chance performance for change localization is lower, even if there is no leak. However, in the Leak model, the gap between the Together- N and Separate conditions is expected to be larger. To make this concrete, we implemented the Leak model by postulating that a fixed proportion of a feature's resource is leaked. Model prediction on average across subjects of the No-leak and the Leak model is shown along with average subject performance in Fig. 5.2. Maximum-likelihood estimates of model parameters are shown in Table 5.1.

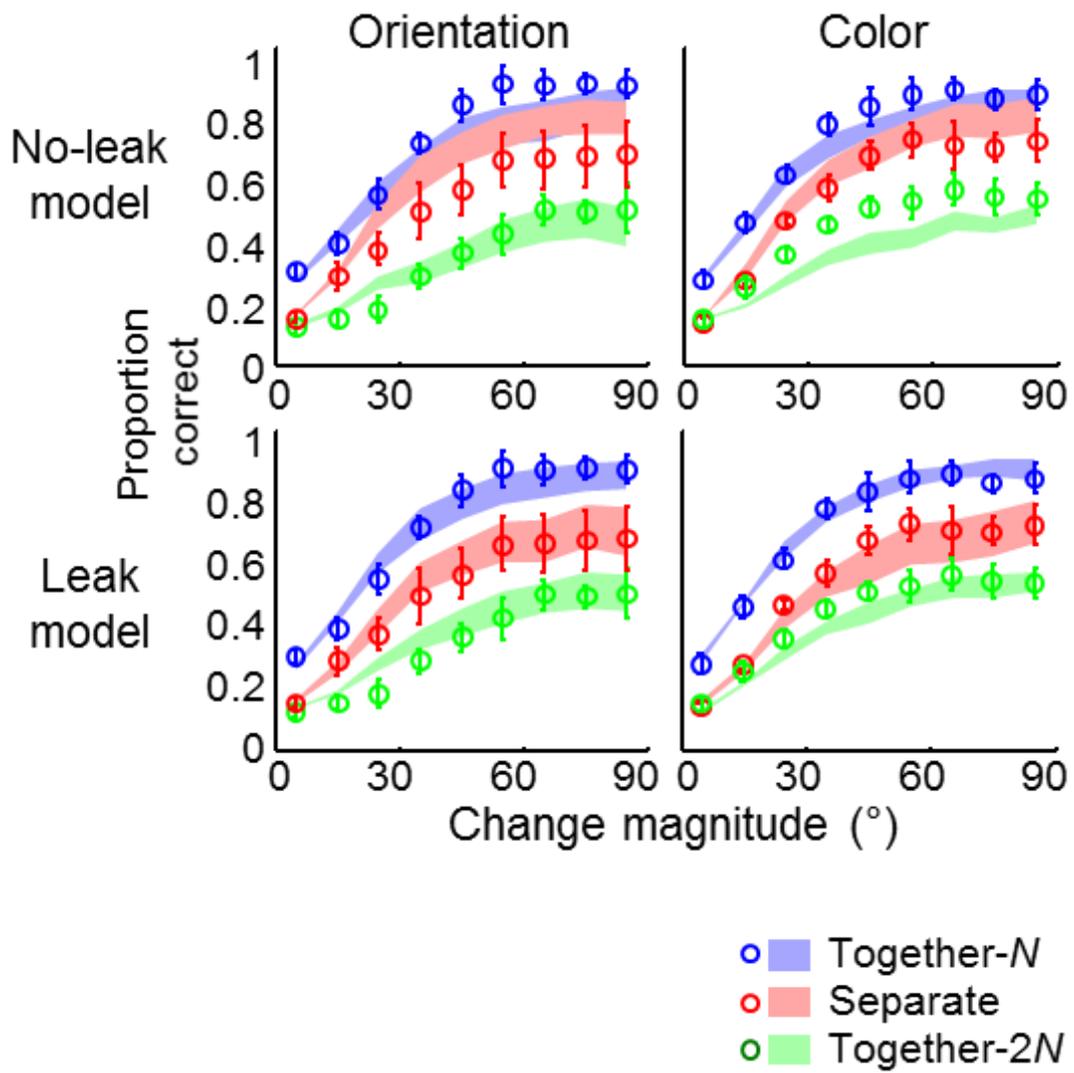


Figure 5.2 Psychometric curves and model predictions. The No-leak model (top) fits better than the Leak model (bottom).

	Parameters					
Models	\bar{J}_{ori}	\bar{J}_{col}	τ_{ori}	τ_{col}	r_{ori}	r_{col}
No-leak	119±17	93±32	157±18	100±6		
Leak	142±12	88±21	157±20	86±19	0.38±0.13	0.210±0.047

Table 5.1 Means and SEs of the maximum-likelihood estimates of model parameters.

4.2 Model comparison

We find that for every individual subject, the AIC of the No-leak model is higher than of the Leak model, on average by 62 ± 17 . We estimate that $38 \pm 13\%$ of orientation resource and $21 \pm 5\%$ of color resource is leaked. However, a model in which we enforce the same leak rate for both features fits nearly equally well (AIC difference: 1.9 ± 1.6) and this leak rate is estimated at $31 \pm 4\%$. The individual AIC values are shown in Table 5.2.

Model	Subject index					Mean
	1	2	3	4	5	
No-leak	3949	4410	3963	4531	3899	4150
Leak	3902	4367	3934	4405	3835	4089

Table 5.2. AIC value of individual subjects for the no-leak and leak models.

5 Discussion

This chapter's finding that memory resource for a particular feature is being exclusively allocated to the objects that have that feature is consistent with a recent delayed-estimation study (41). Here, the authors used orientation and color as well, and as well as with classic change detection studies (66, 72). However, only Olson and Jiang (66) included a Together- $2N$ condition. Even though the authors used highly discriminable feature values, their results are similar to the findings in this chapter: performance was better in a Together- N condition than a Separate condition, and better in the Separate condition than a Together- $2N$ condition.

The difference in performance between the Separate and Together- $2N$ conditions indicates that Separate is not simply a special case of a condition with double the set size as in Together- N . This confirms that in the Separate condition, although feature memory resource leaks, it is still far from being equally allocated to all objects regardless of relevance.

Findings from Chapters 4 and 5 can be revisited by applying other VSTM models than the VP model to the VSTM of two-feature objects by using the idea of factorial model comparison. How a feature is encoded is an independent issue from how two features interplay in multi-feature objects. Thus, it is possible to use, for instance the SA or EP model, to examine the VSTM for the two-features.

Chapter 6

The temporal decay of VSTM

1 Introduction

That VSTM is represented as sustained neural activity during a delay period is well established in physiological studies (16–19, 21). In parallel with these findings, behavioral studies found that observers' performance gradually decreased with duration of delay (27, 79–82), suggesting that neural activity dissipates over time.

Many studies found that VSTM gradually decays (27, 79–81). Earlier studies increased the inter-stimuli interval (ISI) and found that subject performance decreased with ISI (27, 30, 79, 80, 83). Cermak (27) conducted a change detection task using free-form shapes, varied the delays (or ISIs) ($T = 1.5, 4, 12, \text{ and } 20 \text{ sec}$), and found that proportion correct gradually decreases with ISI. Even with the 20-sec delay, performance was substantially above chance. However, Phillips and Baddeley (84) found that with a longer delay ($>10 \text{ sec}$) performance could be as low as chance level.

However, a recent study (82) argued that the number of remembered items decreased without compromising the quality of the precision of the remember items. The authors conducted delayed-estimation experiments (color and shape) with varying durations of delay ($T = 1, 4, \text{ and } 10 \text{ sec}$). They found that the number of remembered items decreased but the precision of remember items did not significantly vary. They interpreted the results as if memory of some items are dropped over time, so called by “sudden death.” However, the problem is that they used the SA model to analyze the data.

In this chapter, I examine the effect of delay duration on VSTM precision in a delayed-estimation task, and find that at all set sizes, subjects are prone to making larger errors when the delay duration is longer. By keeping the consistency, I used a continuous-

resource model, the VP model, to analyze the data. The results suggest that VSTM precision tends to be more variable as the duration of delay increases.

2 Summary of task

I use delayed estimation paradigm in this chapter. In a delayed estimation task, observers' response (estimate of a stimulus) can be hypothesized as a proxy of the stimulus. In the experiments, observers were shown a brief visual display with N oriented ellipses. After a delay of 1 second, observers reported the estimated orientation of one designated item ("target") from the display. The orientations of the ellipses were drawn independently from a uniform distribution. Each item had the same probability of being a target.

3 Theory and models

3.1 Variable-precision model

I have consistently used the VP model to explain experimental results from various VSTM tasks. In this chapter, I use the same VP model again (see Chapter 2 for details). Precision is variable across items and trials.

The error distribution corresponding to a fixed precision J is $p(\hat{s} | s; J) = \text{VM}(\hat{s}; s, \Phi(J))$, where VM represents a Von Mises distribution. When precision is variable, the error distribution is a mixture of the error distributions associated with individual values of precision, with mixture proportions equal to the frequencies of those values, $p(J | \bar{J}; \tau) = \text{Gamma}(J; \bar{J}, \tau)$, with \bar{J} given by Eq.(2.1). Therefore,

$$\begin{aligned}
p(\hat{s}|s; \bar{J}, \tau) &= \int p(s|s; J) p(J|\bar{J}; \tau) dJ \\
&= \int \text{VM}(\hat{s}; s, \Phi(J)) \text{Gamma}(J; \bar{J}, \tau) dJ,
\end{aligned}$$

This is a mixture of an infinite set of Von Mises distributions. We approximate the mixture by sampling 500 values of J from the gamma distribution, and averaging the Von Mises distributions corresponding to these samples. The VP model has three free parameters: \bar{J} , α , and τ .

4 Methods

4.1 Psychophysics

The stimuli were the same as in Chapter 1. Subjects ($n = 5$) viewed a brief display (100 ms) of N gray ellipses ($N = 1, 2, 4, \text{ or } 6$) and memorized the orientation of the ellipses. The orientation of each ellipse was independently drawn from a uniform distribution. After a delay ($T = 1, 2, 3, \text{ or } 6$ sec), subjects reported the orientation of a randomly chosen ellipse. The experiment consisted of four sessions, each run on separate days. In each session, a subject experienced all four durations of delay, which was blocked. Each block consisted of 120 trials and set size varied. Thus, each subject completed $4 \cdot 4 \cdot 120 = 1920$ trials in total.

5 Results

5.1 Estimate error distribution

The histograms of estimation errors of an individual subject are shown in Fig. 6.1. To characterize the distribution of estimation errors, I first computed the circular standard

deviation, which is a measure of spread of directional data. First, a data point (an estimate of error) is represented as a complex unit vector with θ , the angle of the error, $z = \cos \theta + i \sin \theta$. The mean resultant vector is $R = \frac{1}{N} \sum_{n=1}^N z_n$, where N is the total number of data points. The circular standard deviation (CSD) is $\sqrt{-2 \log R}$.

I computed the CSD in radians for a subject's estimation error distribution in each combination of set size and delay (4 set sizes \cdot 4 delay types = 16 in total). A two-way repeated-measures ANOVA reveals a significant main effect of set size ($F(3, 12) = 131.0$, $p < 0.001$), a significant main effect of delay ($F(3, 12) = 7.7$, $p < 0.001$), and no significant interaction ($F(9, 36) = 0.55$, $p = 0.81$). On average, the CSD increases with set size and duration of delay (Fig. 6.2). Mean CSD values across subjects for each set size and delay type are shown in Table 1.

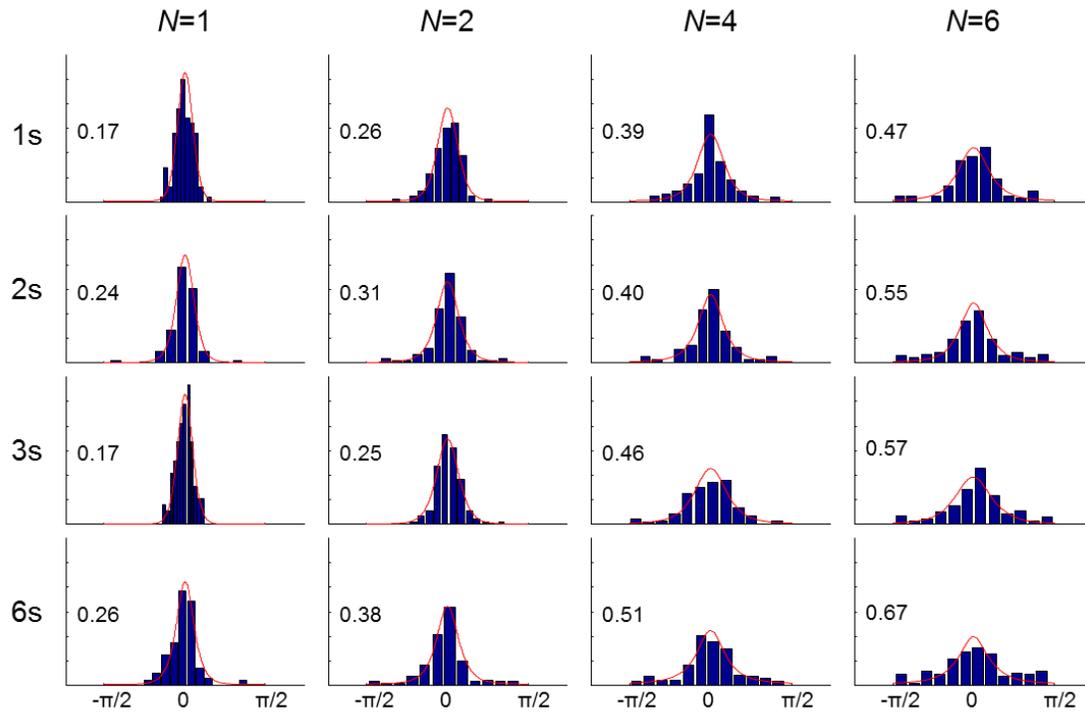


Figure 6.1 Histograms of errors of an individual subject; set size (columns) and duration of delay (rows). Circular standard distribution in radians is shown for each condition. The red line shows the model fits of the variable-precision (VP) model.

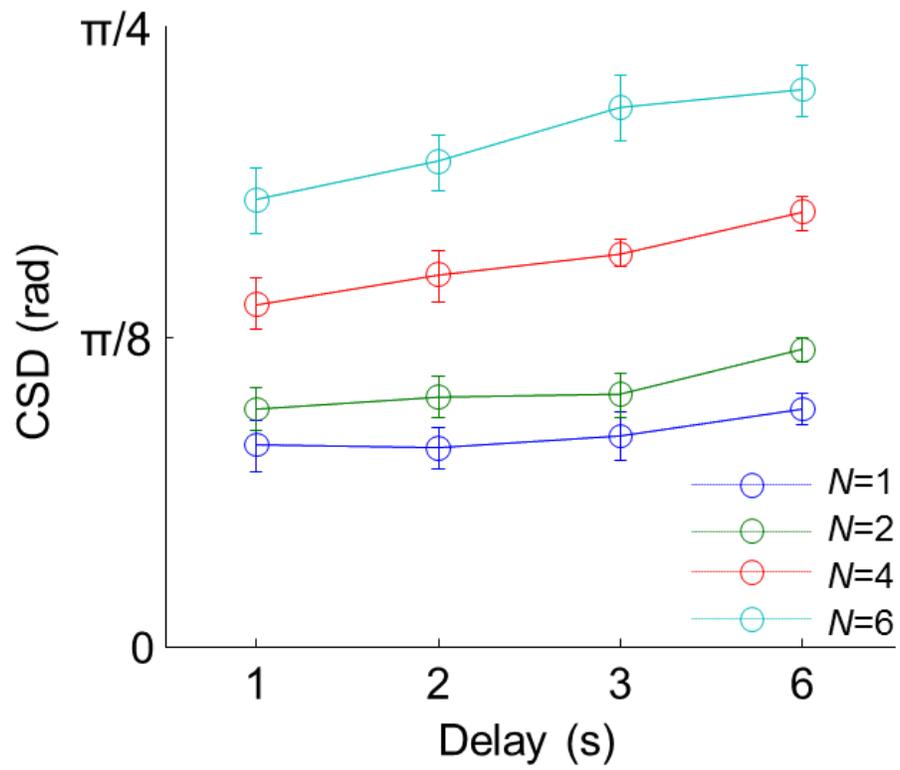


Figure 6.2 Average CSD across subjects for each set size and duration of delay in radians.

Set size	Duration of delay (s)			
	1	2	3	6
1	0.253±0.032	0.250±0.027	0.267±0.029	0.301±0.018
2	0.301±0.027	0.316±0.026	0.318±0.029	0.377±0.016
4	0.433±0.033	0.469±0.031	0.498±0.016	0.549±0.021
6	0.565±0.041	0.613±0.035	0.682±0.042	0.703±0.031

Table 6.1 Mean and s.e.m. of circular standard deviation across subjects from each set size and duration of delay combination.

5.2 Maximum-likelihood estimates of model parameters

Mean and s.e.m. values of maximum-likelihood estimates of a parameter of the VP model are shown in Fig. 6.3. To evaluate the dispersion of precision values of the Gamma distribution, using the parameter estimates from the VP model, I computed the coefficient of variation (CV; standard deviation divided by mean). In the case of a Gamma distribution,

this can be computed from the scale parameter and the mean as $CV = \sqrt{\frac{\tau}{J_1}}$. The CV shows

an increasing trend as a function of delay duration (Fig. 6.3B). To evaluate the CV values across the delay conditions, I conducted a one-way repeated-measure ANOVA. The analysis did not show a significant effect of delay duration ($F(3,4) = 2.78, p = 0.087$).

Duration of delay (s)	J	α	τ
1	28.0±8.1	0.702±0.099	6.8±3.2
2	28.2±5.5	0.80±0.11	6.8±1.8
3	26.2±5.6	0.76±0.14	8.0±1.6
6	21.4±3.1	0.649±0.042	10.2±2.7

Table 6.2 Mean and s.e.m. of maximum-likelihood estimates of the parameters in the VP model.

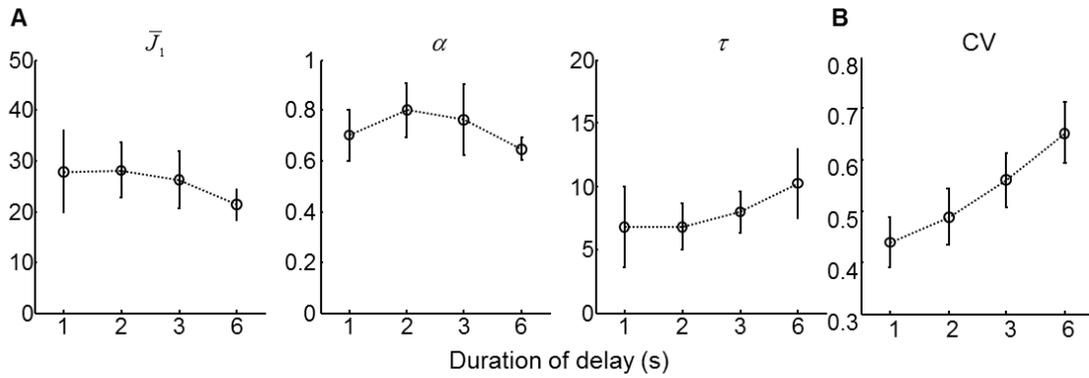


Figure 6.3 (A) Mean and s.e.m. of maximum-likelihood estimates (MLEs) of parameters of the VP model. Left: mean precision (\bar{J}_1), middle: power parameter (α), right: scale parameter (τ). (B) Coefficient of variation (CV) of a Gamma distribution from the VP

model ($CV = \sqrt{\frac{\tau}{J_1}}$).

6 Discussion

The decay of VSTM over time during a delay period is closely related to dissipating sustained neural activity during the delay. However, compared to studies on the set-size effect of VSTM, fewer attempts have been made to model the decay of VSTM. Here, I use a continuous-resource model to examine the decay of VSTM by varying ISIs.

I conducted delayed estimation experiments where subjects memorized stimuli, and reported an estimate of the memorized stimuli after a delay. I examined the effect of varying the durations of the delay on the precision of memory. I found that the distribution of errors increase with the duration of the delay as well as the number of the items presented.

The coefficient of variation based on the maximum-likelihood estimates of the parameters of the VP model did not show any significant difference across delays. This may suggest that even though subject performance drops, the dispersion of precision values does not vary much. In the future, it will be important to consider the time course of VSTM.

Chapter 7

Summary and conclusion

The experiments and models reported in this thesis contribute to a better understanding of the limitations of human VSTM. I proposed a new noise-based model for the limitations of VSTM and rigorously compared the predictions of this model with those of previous models. Examining more complex stimuli that consisted of more than one feature, I established a new noise-based framework that reconciles earlier studies. Finally, I examined the effects of time on the precision of VSTM.

1 Variable precision of visual short-term memory

1.1 Do slots exist?

The results from Chapter 2 suggest that VSTM limitations should be conceptualized in terms of quality of encoding rather than number of items. Earlier work proposing continuous-resource models in the study of VSTM (6-8) did not model variability in resource across items and trials. Here, I have shown that when such variability is not modeled, as in the equal-precision (EP) model, human responses in change localization cannot be accounted for. By contrast, the variable-precision (VP) model accounts for all presented data, including the existence of apparent guessing and its increase with set size, which have so far been attributed to an item limit. Thus, the VP model can be regarded as the first viable continuous-resource model and poses a serious challenge to the notion of slots in VSTM. It might reconcile an apparent capacity of about four items with the subjective sense that one possesses some memory of an entire scene: items are never discarded completely, but their encoding quality could by chance be very low if they receive very little resource.

Most neuroimaging and EEG studies of VSTM limitations only consider the slots framework (5, 21-25) (but see (26-27)). Without testing alternative models of VSTM, these studies cannot provide evidence for the existence of slots. The VP model offers a viable alternative, and I expect that quantities in the VP model will also correlate with neural variables.

1.2 Is resource discrete?

The slots-plus-averaging (SA) model asserts not only that VSTM consists of slots, but also that resource comes in discrete chunks. The latter notion is difficult to reconcile with the fact that sensory noise is a graded rather than a discrete quantity. For example, stimulus contrast affects sensory noise and therefore encoding precision in a graded manner. Such continuous modulation is inconsistent with the allocation of “fixed-size, pre-packaged boxes” (42) of resource, because those only allow for a small, discrete number of noise levels. The VP model does not have this problem, since precision is a continuous quantity and is modulated by contrast in a continuous manner.

1.3 Exploration of further model dimensions

In Chapter 3, variants of both the VP model and previous models were conceived and tested. I generated hybrids such as SA with trial-to-trial variability in capacity K , and VP augmented with an item limit (continuous resource in discrete slots). For rigorousness and consistency across tasks, in addition to change localization, I used change detection data for the factorial model comparison. The results show that the VP model families outperform models consistently across the two tasks as well as for different types of

features (orientation and color). The original VP model used in Chapter 1 was among the winning models in all experiments. The majority of the discrete-capacity models gave worse performance although in case of the slots-plus-averaging (SA) models, when they had variability in the number of encoded items (either uniform- or Poisson-variability), model prediction was much better than when having fixed capacity. Rank correlation analysis revealed that the consistency in model ranking was consistent across tasks and features.

A similar but more thorough meta-analysis, factorial model comparison, shows that the variants of the VP model outperform other models in delayed estimation tasks(54). The last study showed that the data from earlier published studies that used a similar delayed estimation task can be better explained by the variants of the VP model. For instance, interestingly, the authors found that subject data from a study (42), which is considered to be irreconcilable to the variable-precision study, in fact could be well explained by the variants of the VP models. As the authors emphasized, this shows critical importance of formal model comparison to prove the validity of a model.

A direction for future work is to add more dimensions to the VP models. First, different types of variability in precision can be considered: the precision may variable across items, trials, or both. Different decision rules may apply: optimal versus e.g., a suboptimal rule in which subjects ignore variations in stimulus reliability. It is also possible to vary the correlation between the precision with which an item in the first display is encoded with the precision of the (spatially) corresponding item in the second display.

Noise at the decision stage can be added to all models. At the decision stage, subjects may or may not be able to use the deterministic decision rule (85). In addition, it may be that subjects respond using a pure guess on a certain proportion of trials; hence, a fixed lapse rate may be implemented as an additional dimension.

Finally, the independence and exchangeability of objects in the stimuli can be substituted by considering the stimuli as high-order textures. For instance, a display with homogenous stimuli (e.g., similar colors) might contribute to higher precision. Furthermore, possibility of observers using high-order textures can be related to the origin of variability in precision.

1.4 Neural basis of VSTM resource

Previous models have not specified a neural correlate of the VSTM resource. Here, I propose to identify VSTM memory resource with the gain (mean amplitude) of the neural population pattern encoding a stimulus. Several arguments support such an identification. First, for Poisson-like populations, gain is proportional to encoding precision (28). Moreover, the energy cost associated with high gain (29) could explain why working memory is limited: as set size grows larger, the energy cost gradually outweighs the benefit of encoding items with high precision. Finally, gain in visual cortical areas is modulated by attention (30-32), and attentional limitations are closely related to working memory ones (8, 33).

1.5 Decrease of mean precision with set size

The VP model predicts that mean precision decreases gradually with increasing set size, and if encoding precision can be identified with neural gain, that gain does as well. Extant physiological evidence is consistent with this prediction. Neuronal responses in lateral intraparietal cortex (LIP), an area associated with spatial attention, are lower to the onset of four rather than two choice targets (34). In the superior colliculus, an area associated with covert attention, firing rates also decrease with the number of choice targets (35).

In both change localization experiments, I found that the mean precision decreases with set size approximately as $1/N$, which would be predicted by models in which the total amount of resource is, on average, independent of set size. However, the VP model fit to delayed-estimation performance (64) had a steeper decline. This shows that the decrease of mean precision with set size is task-dependent and that the trial-averaged total amount of resource might depend on set size. Perhaps, the precise relation between mean precision and set size is set by a trade-off between energy expenditure and performance. In support of this speculation, a decrease of mean precision with set size is also observed in an attentionally demanding task without a memory component (36).

2 Visual short-term memory for more complex objects

Many behavioral studies use simple stimuli such as colored discs, oriented ellipses, or Gabor patches to study VSTM, but the ultimate goal is to understand how VSTM operates within rich and complex human surroundings. In order to answer this question, it is necessary to take a step forward from simple stimuli without making the stimuli too

complex to measure. In Chapters 4 and 5, I made such an attempt by using objects with two features, orientation and color.

However, I found that earlier studies on VSTM for multi-feature objects have been mostly based on the discrete-capacity view, framed as a dichotomous question: whether the basic unit of VSTM is an object or a feature. This question has always been roughly into three types of hypotheses (Tab. 7.1). However, they become partially uninterpretable when memory noise is considered, and fail to explain the complexity of the data from various experimental conditions. Therefore, a less dichotomous view of multi-feature VSTM is needed.

The discrete-capacity view		The continuous-resource view	
H1	If VSTM is object-based, then the memory of a feature should not suffer from the addition of a second feature to the same object.	Resource allocation among features	Q1
H2	If VSTM is object-based, then encoding a task-relevant feature of an object should automatically cause irrelevant features of that object to be encoded as well.	Encoding and role in decision-making of irrelevant features.	Q2
H3	If VSTM is object-based, then memorizing two features of the same object should be easier than of two different objects.	Spatial allocation of resource	Q3

Table 7.1 Comparison of the previously posed hypotheses (H1-H3) on the VSTM for multi-feature objects, which were based on the discrete-capacity views, and the new questions (Q1-Q3) reconceptualized by using the continuous-resource view.

2.1 New framework for VSTM of orientation and color

Here, I propose a conceptually related but different framework (Fig. 7.1A). Each feature has an independent pool of VSTM resources. Some of the resource for a given feature gets placed into multi-feature packages, which are then distributed across all objects, while the rest gets allocated in a targeted manner to only the objects for which that feature is relevant.

Brady et al. (86) have suggested that the unit of VSTM is a “hierarchically structured feature bundle”. The hierarchy consists of two levels: at the bottom level, each feature can be stored or forgotten independently. At the top level, they are integrated into a “bundle”. I use the term “package” for resources, in contrast to Brady et al.’s term “bundle” for features. Within the multi-feature package, resource remains feature-specific, and if the object does not have one of the features, the corresponding resource from the package gets wasted. Finally, the brain utilizes a smart decoder, in the sense that the signal arising from resources being allocated to an irrelevant feature can be ignored during decision-making.

This framework can account for all our findings as well as those of others (41, 75) (Fig. 7.1B). In Experiments 1 and 2, the amounts of packaged and targeted resource received by each feature of each object are the same across conditions, explaining the identical performance. Experiment 3 establishes the smartness of the decoder. In Experiment 4, the amount of packaged resource is the same regardless of whether a feature is relevant or irrelevant, but in the latter case, the feature does not receive the targeted resource and performance is lower. In Experiment 5, precision in the “Separate” condition is lower than in the “Together- N ” condition because the *packaged* resource is distributed over double the number of objects, even though targeted resource is distributed over the

same number of objects. This also explains the results of (41, 66, 72) and corresponds to an “object benefit” (37) for the “Together- N ” condition. Moreover, precision in “Separate” is higher than in the “Together- $2N$ ” condition because the *targeted* resource is distributed over half the number of objects, even though packaged resource is distributed over the same number of objects. This is consistent with one comparison in (75) and with classic change detection results (66). However, it is inconsistent with a comparison in (75) that used a condition in which two sequentially presented displays differed by which feature was relevant for memorization (for example, color had to be remembered for the stimuli in the first display, orientation for those in the second). Performance was lower when the stimuli possessed an irrelevant feature in addition to their relevant feature; the discrepancy with our framework might be due to the fact that relevance switched mid-trial, causing targeted resource to be allocated to an irrelevant feature. Although more tests are needed, the packaged/targeted framework seems a promising candidate for a general theory of multi-item VSTM with noisy memories.

2.2 VSTM for objects with other feature combinations

One direction for further study is to test different feature combinations. There are indications that some combinations behave the same way as orientation + color, and others differently. For example, a study of H1 for color-shape combinations reported that performance was lower when subjects memorized both features than when they memorized only one (33), although contradictory results were found later (50, 87). Results on combinations of size, orientation, and color have been similarly mixed (28, 88). A recent study using orientation, color, shape, spatial frequency, and size found a decrease of

performance when the number of features increases in an object (89), but the different pairs were not analyzed separately. Regarding H2, irrelevant changes in orientation might impair location memory (90), and the effect of irrelevant shape on color memory seems to be predominant at higher set sizes (68, 71). Regarding H3, using combinations of size, orientation, and shape, performance was found to be similar between the Together- N and the Separate condition (72). Although these differences between feature combinations are confusing, we believe that the first step in all cases is to measure full psychometric curves in a battery of standardized experiments similar to the one used here. Finally, as soon as one tries to scale studies of VSTM up to more naturalistic objects, it becomes unclear what their elementary features are, and a fundamentally different approach might be called for.

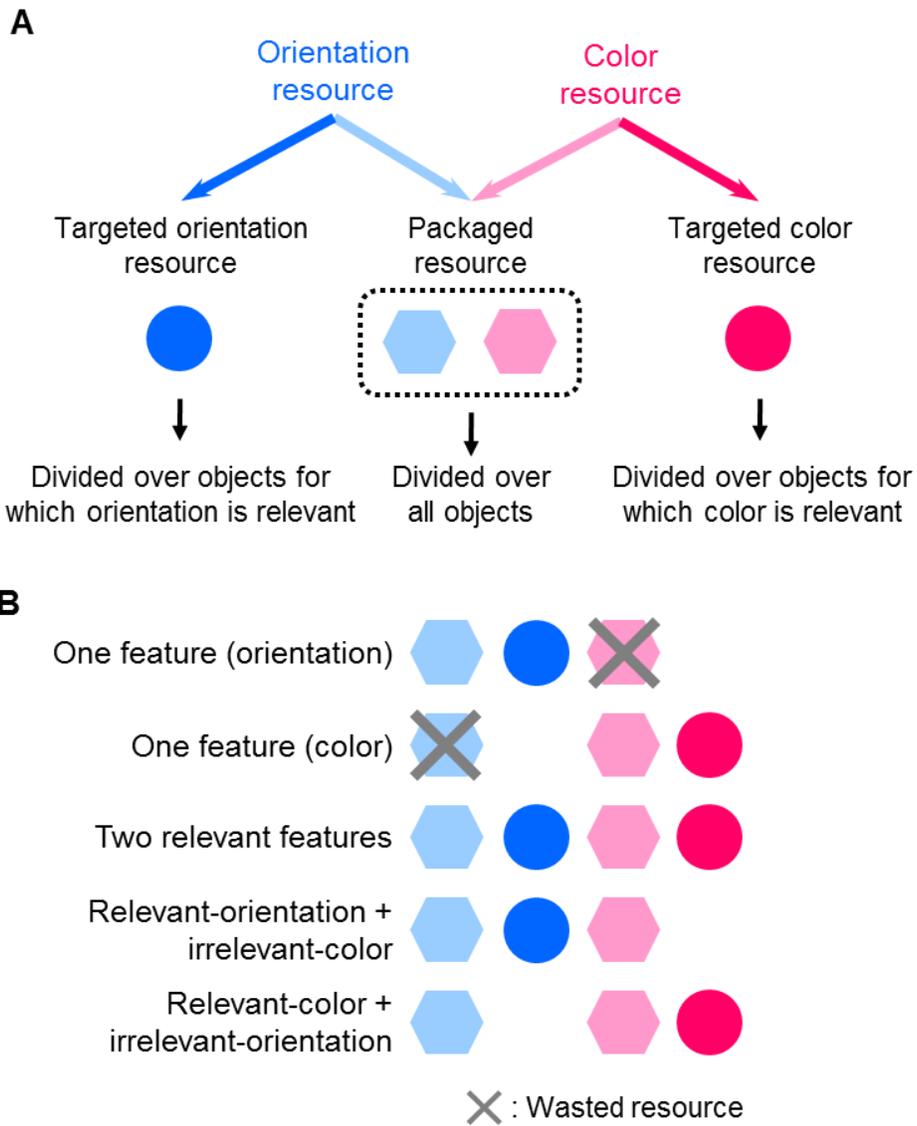


Figure 7.1 New framework for VSTM of orientation and color. (A) Feature resource consists of “packaged” resource, which gets allocated to all objects, and “targeted” resource, which only gets allocated to objects for which that feature is relevant. (B) Application to the conditions from Chapter 4.

2.3 Definition of an object

A major criticism on the feature integration theory (67) is that there is no clear definition of what constitutes a feature (91). Similarly, the definition of an object is a subject of discussion, and Feldman (92) argues that perception of a visual object rather depends on the observer's interpretation of the hierarchical organization of a visual scene, rather than the physical properties of the scene.

In fact, in earlier studies on VSTM for multiple features, the definition of an object or a feature has been overlooked. Luck and Vogel (28) found that performance in a change detection task using bi-colored objects (e.g., a smaller square of color A superimposed on a larger square of color B, where A and B are highly distinctive colors) was not significantly different from single-color objects. However, Wheeler and Treisman (33) found that performance decreased with the bi-color stimuli. It is questionable whether subjects truly perceived a bi-colored stimulus as a unitary object. Similarly, studies mentioned in Chapter 5, which examined the effect of the total number of features on performance (32, 41, 72), used stimuli consisted of two items where one was on top of the other.

To minimize the ambiguity, in Chapters 4 and 5, I used stimuli with the two features, where each feature had only one value in a stimulus; each stimulus had one orientation value and one color value. In addition, all objects had homogenous texture, and had one continuous boundary with the background.

2.4 Location as a feature

Location and identity of an object are two separate pieces of information. In general, studies have reached a consensus on the fact that memory of location and identity is stored separately (93–95). The binding of location and identity has been a major topic of VSTM literature. Researchers studied erroneous report of identity of an object where subjects mistakenly report the identity of an object by reporting the identity of a different object of the display. They are called conjunction errors, illusory conjunctions, or swap errors (33, 96). To avoid possible complications from manipulating location and identity of an object at the same time, in Chapters 4 and 5, the locations of stimuli did not vary within a condition.

One possible way to test whether observers swap the identity of a stimulus from another is to build a hierarchical inference model. A simple design of an experiment can be a delayed estimation task where observers first view a set of items followed by a delay, then view a location or an identity probe, and estimate the other feature (either the identity or the location). Location can be assumed to be an independent feature that is encoded with noise. Location and identity can be bound in a higher layer with noise.

2.5 Color as an ecological relevant feature

In Experiment in Chapter 4, I found that the memory of an irrelevant feature was better in the color domain than in the orientation domain. There are several explanations for this finding. First, differences in response modality might have brought different response times. Subjects saw a color wheel, which has the entire spectrum of all possible colors in the color trials, but for the orientation trials, subjects only saw one orientation at a time because they moved a mouse to change the orientation of the response probe. Therefore, this might have

resulted in the subjects responding with a longer delay in the orientation trials, which could trigger lower precision of memory.

In nature, colors are known to represent the identity of an object (97–99) such as warning signs from poisonous prey or the presence of a predator. This implies that color tends to convey more salient information than orientation in nature. Thus, color VSTM may have a slower decay compared to orientation.

3 Temporal dynamics of VSTM

In Chapter 6, I showed that delayed estimation data can be fitted to the VP model.

One can assume an exponential decay of precision over time. This precision decay can occur at the level of the mean precision, or independently at the level of the precision of each individual stimulus. If it is the latter case, it is possible to implement independent decay of each item over time. Similarly, one can implement the “sudden death” model (82) by assuming that after a certain duration of time, a certain number of items are dropped from memory with an equal probability.

It is possible to use all the models that I introduced in Chapter 3 and add temporal dimension to all models to conduct another factorial model comparison. More modeling dimensions can be added. In change localization and change detection, a trial consists of two displays. One can assume independent decay patterns across displays.

Another direction is to compare the pattern of decay across different types of features. Experiments for testing the decay of relevant feature memory is possible, but conducting an experiment that is similar to Experiment 4 of Chapter 4 (Amazon Mechanical Turk) is possible. In this case, two questions can be raised; 1) whether

irrelevant feature memory is more vulnerable to decay, and 2) if yes, and if VSTM decay is more manifested in irrelevant feature memory, whether there is difference in the pattern of decay across features.

4 Conclusion

Overall, the findings in this thesis pose a serious challenge to the prevailing discrete-capacity view of VSTM limitations and support a new alternative model, the variable-precision model, in which memory is conceptualized as a continuous resource with variability. To examine the new model, I implemented previously suggested VSTM models, and formally compared the predictions of various models accompanied by thorough examination of each model factor. After establishing a rigorous foundation for the variable-precision model, I applied this new framework to investigate other issues in VSTM studies that had previously been discussed based on the discrete-capacity view. First, as a step forward to the ultimate goal of understanding VSTM for natural stimuli, I investigated the VSTM for objects with multiple features. I reinterpreted previously suggested hypotheses into new questions using the continuous-resource view and conducted a battery of experiments. Based on the results, I propose a new framework, in which orientation and color has an independent memory resource, but some portion of each resource is packaged together and distributed to all objects. This view not only explains all the results from this thesis but also reconciles the debate across most of the previous findings. Finally, I used the same variable-precision model to explore the temporal aspects of VSTM. This body of work provides a guideline to examine various aspects of human VSTM in a more rigorous

way, contributes to a better understanding on the limitations of human VSTM, and suggests that it is time to reconsider the discrete-capacity view on VSTM limitations.

Bibliography

1. Gibson JJ (2013) *The Ecological Approach To Visual Perception* (Psychology Press).
2. Baddeley A (2003) Working memory: looking back and looking forward. *Nat Rev Neurosci* 4(10):829–839.
3. Rensink RA (2002) Change detection. *Annu Rev Psychol* 53(1):245–277.
4. Irwin DE (1991) Information integration across saccadic eye movements. *Cognit Psychol* 23(3):420–456.
5. Prime SL, Niemeier M, Crawford JD (2005) Transsaccadic integration of visual features in a line intersection task. *Exp Brain Res* 169(4):532–548.
6. Hollingworth A, Richard AM, Luck SJ (2008) Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *J Exp Psychol Gen* 137(1):163–181.
7. Rayner K (1998) Eye movements in reading and information processing: 20 years of research. *Psychol Bull*:372–422.
8. O'Regan JK, Le'vy-Schoen A (1983) Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Res* 23(8):765–768.
9. Henderson JM, Hollingworth A (1999) The Role of Fixation Position in Detecting Scene Changes Across Saccades. *Psychol Sci* 10(5):438–443.

10. Brouwer A-M, Knill DC (2007) The role of memory in visually guided reaching. *J Vis* 7(5):6.
11. Cattell JM (1886) The time it takes to see and name objects. *Mind*.
12. Kaufman EL, Lord MW (1949) The discrimination of visual number. *Am J Psychol* 62(4):498–525.
13. Pylyshyn ZW, Storm RW (1988) Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat Vis* 3(3):179–197.
14. Pylyshyn Z (2004) Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Vis Cogn* 11(7):801–822.
15. Yantis S (1992) Multielement visual tracking: attention and perceptual organization. *Cognit Psychol* 24(3):295–340.
16. Fuster JM, Alexander GE (1971) Neuron Activity Related to Short-Term Memory. *Science* 173(3997):652–654.
17. Kubota K, Niki H (1971) Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J Neurophysiol* 34(3):337–347.
18. Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol* 61(2):331–349.
19. Courtney SM, Ungerleider LG, Keil K, Haxby JV (1997) Transient and sustained activity in a distributed neural system for human working memory. *Nature* 386(6625):608–611.
20. Goldman-Rakic PS (1995) Cellular basis of working memory. *Neuron* 14(3):477–485.

21. Curtis CE, D'Esposito M (2003) Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci* 7(9):415–423.
22. Schluppeck D, Curtis CE, Glimcher PW, Heeger DJ (2006) Sustained activity in topographic areas of human posterior parietal cortex during memory-guided saccades. *J Neurosci Off J Soc Neurosci* 26(19):5098–5108.
23. Fuster JM (2001) The Prefrontal Cortex—An Update: Time Is of the Essence. *Neuron* 30(2):319–333.
24. Hebb DO (1949) *The organization of behavior* (New York: Wiley).
25. Rensink RA, O'Regan JK, Clark JJ (1997) To See or not to See: The Need for Attention to Perceive Changes in Scenes. *Psychol Sci* 8(5):368–373.
26. Hollingworth A, Henderson JM (1999) Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination. *Acta Psychol (Amst)* 102(2-3):319–343.
27. Cermak GW (1971) Short-term recognition memory for complex free-form figures. *Psychon Sci*:209–211.
28. Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390(6657):279–281.
29. Palmer J (1990) Attentional limits on the perception and memory of visual information. *J Exp Psychol Hum Percept Perform* 16(2):332.
30. Phillips WA (1974) On the distinction between sensory storage and short-term visual memory. *Percept Psychophys* 16(2):283–290.

31. Vogel EK, Woodman GF, Luck SJ (2001) Storage of features, conjunctions, and objects in visual working memory. *J Exp Psychol Hum Percept Perform* 27(1):92–114.
32. Xu Y (2002) Limitations of object-based feature encoding in visual short-term memory. *J Exp Psychol Hum Percept Perform* 28(2):458–468.
33. Wheeler ME, Treisman AM (2002) Binding in short-term visual memory. *J Exp Psychol Gen* 131(1):48–64.
34. Alvarez GA, Cavanagh P (2004) The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol Sci* 15(2):106–111.
35. Fougne D, Suchow JW, Alvarez GA (2012) Variability in the quality of visual working memory. *Nat Commun* 3:1229.
36. Keshvari S, van den Berg R, Ma WJ (2013) No evidence for an item limit in change detection. *PLoS Comput Biol* 9(2):e1002927.
37. Jiang Y, Olson IR, Chun MM (2000) Organization of visual short-term memory. *J Exp Psychol Learn Mem Cogn* 26(3):683–702.
38. Pashler H (1988) Familiarity and visual change detection. *Percept Psychophys* 44(4):369–378.
39. Wilken P, Ma WJ (2004) A detection theory account of change detection. *J Vis* 4(12):1120–1135.
40. Bays PM, Husain M (2008) Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science* 321(5890):851–854.

41. Fougne D, Asplund CL, Marois R (2010) What are the units of storage in visual working memory? *J Vis* 10(12):27, 1–11.
42. Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. *Nature* 453(7192):233–235.
43. Buschman TJ, Siegel M, Roy JE, Miller EK (2011) Neural substrates of cognitive capacity limitations. *Proc Natl Acad Sci* 108(27):11252–11255.
44. Johnson MK, et al. (2013) The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia. *Neuropsychology* 27(2):220–229.
45. Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 63(2):81–97.
46. Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 24(1):87–114; discussion 114–185.
47. Cowan N, Blume CL, Saults JS (2013) Attention to attributes and objects in working memory. *J Exp Psychol Learn Mem Cogn* 39(3):731–747.
48. Anderson DE, Vogel EK, Awh E (2011) Precision in Visual Working Memory Reaches a Stable Plateau When Individual Item Limits Are Exceeded. *J Neurosci* 31(3):1128–1138.
49. Awh E, Barton B, Vogel EK (2007) Visual working memory represents a fixed number of items regardless of complexity. *Psychol Sci* 18(7):622–628.
50. Luria R, Vogel EK (2011) Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia* 49(6):1632–1639.

51. Ma WJ, Husain M, Bays PM (2014) Changing concepts of working memory. *Nat Neurosci* 17(3):347–356.
52. Pouget A, Dayan P, Zemel R (2000) Information processing with population codes. *Nat Rev Neurosci* 1(2):125–132.
53. Shaw ML (1980) Identifying attentional and decision-making components in information processing.
54. Van den Berg R, Awh E, Ma WJ (2014) Factorial comparison of working memory models. *Psychol Rev* 121(1):124–149.
55. Ma WJ, Huang W (2009) No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint. *J Vis* 9(11):3–3.
56. Nienborg H, Cumming BG (2009) Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature* 459(7243):89–92.
57. Cohen MR, Maunsell JHR (2011) Using Neuronal Populations to Study the Mechanisms Underlying Spatial and Feature Attention. *Neuron* 70(6):1192–1204.
58. Lara AH, Wallis JD (2012) Capacity and precision in an animal model of visual short-term memory. *J Vis* 12(3):13, 1–12.
59. Lara AH, Wallis JD (2014) Executive control processes underlying multi-item working memory. *Nat Neurosci* 17(6):876–883.
60. Tanner Jr. WP, Swets JA (1954) A decision-making theory of visual detection. *Psychol Rev* 61(6):401–409.
61. Abramowitz M, Stegun IA (1972) *Handbook of Mathematical Functions* (Dover, New York).

62. Cover TM, Thomas JA (1991) *Elements of Information Theory* (Wiley, New York).
63. Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9(11):1432–1438.
64. Van den Berg R, Shin H, Chou W-C, George R, Ma WJ (2012) Variability in encoding precision accounts for visual short-term memory limitations. *Proc Natl Acad Sci* 109(22):8780–8785.
65. Jeffreys H (1961) *The Theory of Probability* (Oxford).
66. Olson IR, Jiang Y (2002) Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Percept Psychophys* 64(7):1055–1067.
67. Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cognit Psychol* 12(1):97–136.
68. Xu Y (2010) The Neural Fate of Task-Irrelevant Features in Object-Based Processing. *J Neurosci* 30(42):14020–14028.
69. Shen M, Tang N, Wu F, Shui R, Gao Z (2013) Robust object-based encoding in visual working memory. *J Vis* 13(2):1, 1–11.
70. Hyun J, Woodman GF, Vogel EK, Hollingworth A, Luck SJ (2009) The comparison of visual working memory representations with perceptual inputs. *J Exp Psychol Hum Percept Perform* 35(4):1140–1160.
71. Yin J, et al. (2012) Does high memory load kick task-irrelevant information out of visual working memory? *Psychon Bull Rev* 19(2):218–224.
72. Lee D, Chun MM (2001) What are the units of visual short-term memory, objects or spatial locations? *Percept Psychophys* 63(2):253–257.

73. Keshvari S, van den Berg R, Ma WJ (2012) Probabilistic Computation in Human Perception under Variability in Encoding Precision. *PLoS ONE* 7(6):e40216.
74. Pearson B, Raskevicius J, Bays PM, Pertzov Y, Husain M (2014) Working memory retrieval as a decision process. *J Vis* 14(2):2–2.
75. Marshall L, Bays PM (2013) Obligatory encoding of task-irrelevant features depletes working memory resources. *J Vis* 13(2):21, 1–13.
76. Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
77. Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis* 10:437–442.
78. Mardia KV, Jupp PE (1999) *Directional statistics* (John Wiley and Sons).
79. Cornelissen FW, Greenlee MW (2000) Visual memory for random block patterns defined by luminance and color contrast. *Vision Res* 40(3):287–299.
80. Lee B, Harris J (1996) Contrast Transfer Characteristics of Visual Short-term Memory. *Vision Res* 36(14):2159–2166.
81. Petrides M, Baddeley A (1996) Specialized Systems for the Processing of Mnemonic Information within the Primate Frontal Cortex [and Discussion]. *Philos Trans R Soc Lond B Biol Sci* 351(1346):1455–1462.
82. Zhang W, Luck SJ (2009) Sudden death and gradual decay in visual working memory. *Psychol Sci* 20(4):423–428.
83. Bennett PJ, Cortese F (1996) Masking of spatial frequency in visual memory depends on distal, not retinal, frequency. *Vision Res* 36(2):233–238.

84. Phillips WA, Baddeley AD (1971) Reaction time and short-term visual memory. *Psychon Sci* 22(2):73–74.
85. Mueller ST, Weidemann CT (2008) Decision noise: An explanation for observed violations of signal detection theory. *Psychon Bull Rev* 15(3):465–494.
86. Brady TF, Konkle T, Alvarez GA (2011) A review of visual memory capacity: Beyond individual items and toward structured representations. *J Vis* 11(5):4–4.
87. Woodman GF, Vogel EK (2008) Selective storage and maintenance of an object's features in visual working memory. *Psychon Bull Rev* 15(1):223–229.
88. Hardman KO, Cowan N (2014) Remembering Complex Objects in Visual Working Memory: Do Capacity Limits Restrict Objects or Features? *J Exp Psychol Learn Mem Cogn*. Available at: <http://doi.apa.org/getdoi.cfm?doi=10.1037/xlm0000031> [Accessed December 10, 2014].
89. Oberauer K, Eichenberger S (2013) Visual working memory declines when more features must be remembered for each object. *Mem Cognit*. Available at: <http://link.springer.com/10.3758/s13421-013-0333-6> [Accessed June 5, 2013].
90. Jiang Y, Chun MM, Olson IR (2004) Perceptual grouping in change detection. *Percept Psychophys* 66(3):446–453.
91. Briand KA, Klein RM (1989) Has feature integration theory come unglued? A reply to Tsal. *J Exp Psychol Hum Percept Perform* 15(2):401–406.
92. Feldman J (2003) What is a visual object? *Trends Cogn Sci* 7(6):252–256.
93. Pertzov Y, Dong MY, Peich M-C, Husain M (2012) Forgetting What Was Where: The Fragility of Object-Location Binding. *PLoS ONE* 7(10):e48214.

94. Kahneman D, Treisman A, Gibbs BJ (1992) The reviewing of object files: Object-specific integration of information. *Cognit Psychol* 24(2):175–219.
95. Xu Y, Chun MM (2009) Selecting and perceiving multiple visual objects. *Trends Cogn Sci* 13(4):167–174.
96. Treisman A, Schmidt H (1982) Illusory conjunctions in the perception of objects. *Cognit Psychol* 14(1):107–141.
97. Tanaka J, Weiskopf D, Williams P (2001) The role of color in high-level vision. *Trends Cogn Sci* 5(5):211–215.
98. Yao AYJ, Einhäuser W (2008) Color aids late but not early stages of rapid natural scene recognition. *J Vis* 8(16):12.
99. Wichmann FA, Sharpe LT, Gegenfurtner KR (2002) The contributions of color to recognition memory for natural scenes. *J Exp Psychol Learn Mem Cogn* 28(3):509–520.