Visual short-term memory for oriented, colored objects

Hongsup Shin

Department of Neuroscience,
Baylor College of Medicine, Houston, TX, USA
Center for Neural Science and Department of Psychology,
New York University, New York, NY, USA

Wei Ji Ma

Department of Neuroscience,
Baylor College of Medicine, Houston, TX, USA
Center for Neural Science and Department of Psychology,
New York University, New York, NY, USA

A central question in the study of visual short-term memory (VSTM) has been whether its basic units are objects or features. Most studies addressing this question have used change detection tasks in which the feature value before the change is highly discriminable from the feature value after the change. This approach assumes that memory noise is negligible, which recent work has shown not to be the case. Here, we investigate VSTM for orientation and color within a noisy-memory framework, using change localization with a variable magnitude of change. A specific consequence of the noise is that it is necessary to model the inference (decision) stage. We find that (a) orientation and color have independent pools of memory resource (consistent with classic results); (b) an irrelevant feature dimension is either encoded but ignored during decision-making, or encoded with low precision and taken into account during decision-making; and (c) total resource available in a given feature dimension is lower in the presence of task-relevant stimuli that are neutral in that feature dimension. We propose a framework in which feature resource comes both in packaged and in targeted form.

Introduction

A classic question in the study of visual short-term memory (VSTM) has been whether multi-feature objects get stored in VSTM as whole objects or as loose sets of features. Over the past two decades, this question has been turned into testable hypotheses in at least three, not mutually exclusive, ways:

Hypothesis 1 (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. In change detection tasks, whether changes occur only in orientation, only in color, or in either does not seem to affect performance (Luck & Vogel, 1997; Olson & Jiang, 2002; Vogel, Woodman, & Luck, 2001). This finding seems to rule out that VSTM stores a fixed total number of features (summed across all objects), but is compatible with VSTM storing *K* objects regardless of their number of features (Luck & Vogel, 1997; Vogel et al., 2001). However, adding a second color feature to a color-defined object does decrease performance (Wheeler & Treisman, 2002). These findings can be reconciled by a model in which each feature dimension has a separate capacity (Olson & Jiang, 2002; Wheeler & Treisman, 2002). More recently, this model has been challenged by effects of number of features in objects with up to six features (Hardman & Cowan, 2015; Oberauer & Eichenberger, 2013).

Hypothesis 2 (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. This hypothesis has been tested by examining whether the addition of an irrelevant feature decreases performance. The finding that it does not (Luria & Vogel, 2011; Shen, Tang, Wu, Shui, & Gao, 2013; Vogel et al., 2001; Xu, 2010; but see Hyun, Woodman, Vogel, Hollingworth, & Luck, 2009; Yin et al., 2012) has been interpreted as evidence that the irrelevant feature dimension is not encoded (Luria & Vogel, 2011; Vogel et al., 2001). However, this interpretation relies on two implicit assumptions: that the irrelevant feature dimension does not have its own capacity, and that if the irrelevant feature dimension were encoded, it would be ignored during decision-making—a problem that to our knowledge has not previously been noted.

Hypothesis 3 (H3): If VSTM is object-based, then remembering two features of the same object should be easier than of two different objects. Such an "object benefit" (Jiang, Olson, & Chun, 2000) has been tested

for by comparing a condition in which N objects each have two features to one in which $2N$ objects each have one (Lee & Chun, 2001; Olson & Jiang, 2002; Xu, 2002). Performance was found to be lower in the latter condition, suggesting that VSTM is weakly object-based (Olson & Jiang, 2002). However, the fact that spatial attention is divided over more objects in the second condition complicates this conclusion.

Across the three hypotheses, most studies designed to test whether VSTM is object-based have employed change detection tasks with highly discriminable stimuli, such as primary and secondary colors. Such stimuli were chosen with the intention of avoiding 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. In recent years, however, it has increasingly been recognized that VSTM encoding noise can be high enough to cause signal detection errors (Bays & Husain, 2008; Keshvari, Van den Berg, & Ma, 2012, 2013; Lara & Wallis, 2012; Van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004) or large estimation errors (Fougnie, Suchow, & Alvarez, 2012; Oberauer & Lin, 2017; Van den Berg, Awh, & Ma, 2014; Van den Berg et al., 2012; Wilken & Ma, 2004; Zhang & Luck, 2008). Whereas the exact nature of memory noise is still subject to debate (Luck & Vogel, 2013; Ma, Husain, & Bays, 2014; Oberauer & Lin, 2017; Sims, 2015), it is by now no longer disputed that a theory of VSTM cannot be complete without taking into account memory noise.

In change detection and related paradigms, a noisy-memory 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 near zero (Van den Berg et al., 2014). Second, retrieval, including decision-making, becomes an active process of probabilistic inference (Keshvari et al., 2012; Pearson, Raskevicius, Bays, Pertzov, & Husain, 2014; Van den Berg et al., 2012): the observer decides which hypothesis—change or no change—is better supported by the noisy stimulus measurements from both displays. Differences between conditions are no longer necessarily due to encoding only.

For these reasons, it is interesting to revisit the hypotheses H1–H3 within the conceptual framework of noisy short-term memories. This framework comes naturally with the concept of memory resource (Bays & Husain, 2008; Ma et al., 2014). More memory resource allocated to a stimulus implies that the measurements of that stimulus have lower noise (therefore, higher precision), at least on average. Using the concepts of noisy memories and resource, we can reformulate the questions corresponding to H1–H3 as follows.

Question 1 (Q1): Resource allocation among feature dimensions. In a noisy-memory view, the analog of the question "Does each feature dimension have a separate capacity?" (Olson & Jiang, 2002; Wheeler & Treisman, 2002) would be "Is memory resource shared among feature dimensions of an object, or does each feature dimension have its own pool of resource?"

Question 2 (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 (Keshvari et al., 2012, 2013; Ma & Huang, 2009; Wilken & Ma, 2004), 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 (see H2 above).

Question 3 (Q3): Spatial allocation of resource. H3 predicts that dividing the features of N two-feature objects over $2N$ one-feature objects decreases performance. Here, we ask by analogy whether dividing the features of N two-feature objects over $2N$ one-feature objects decreases the amount of resource in a given feature dimension that is available to encode the N feature values in that dimension. In addressing this question, we will account in the analysis for the main effect that localizing a target (here: a change) among $2N$ items is intrinsically harder than among N items.

Aside from its conceptual framework, our study differs from most of the previous literature on object-based VSTM in the following ways:

1. We use the relatively rare paradigm of *change localization*. In change detection, chance level is 0.5. In change localization with N items, chance level is $1/N$. This affords a larger performance range, potentially allowing the predictions of different models to separate more. We have previously used change localization to distinguish one-feature VSTM encoding models (Van den Berg et al., 2012).

2. We vary the magnitude of change. Doing so allows for a precise description of the role of memory noise: more noise means a shallower psychometric curve over change magnitude (Bays & Husain, 2008; Keshvari et al., 2012, 2013; Lara & Wallis, 2012; Pearson et al., 2014; Van den Berg et al., 2012). Using highly discriminable stimuli across a change 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
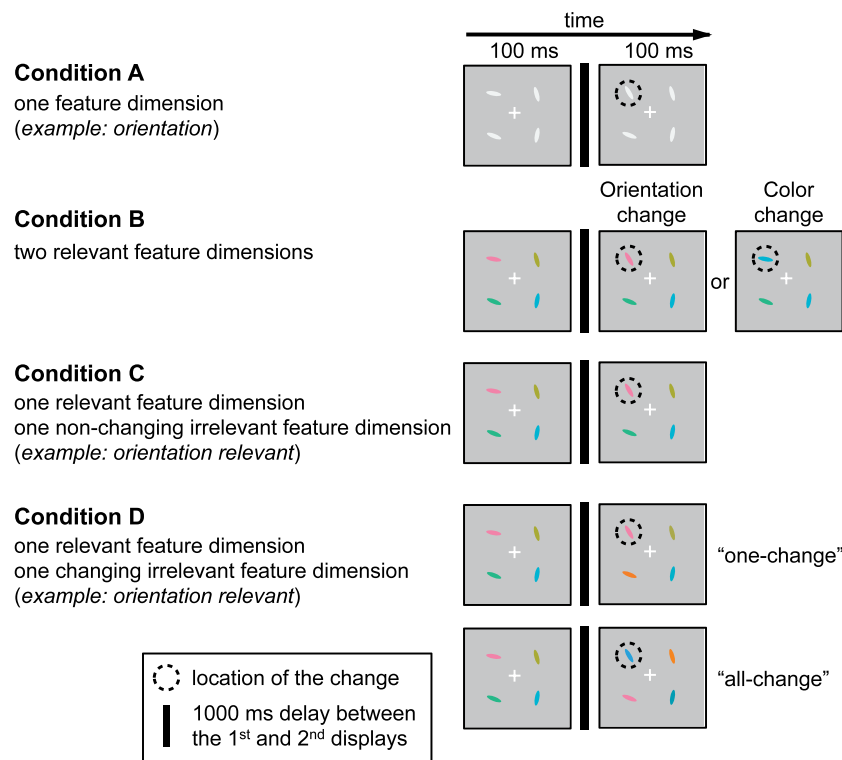
Figure 1. Conditions in Experiments 1, 2, and 3. Subjects click on the location where a (relevant) change has occurred. (A) Orientation trial in Condition A: orientation is the only feature dimension. (B) Condition B: stimuli have two feature dimensions and the change occurs in either. (C) Orientation trial in Condition C: stimuli have two feature dimensions but only orientation is relevant. No changes occur in the irrelevant feature dimension (in this case, color). (D) Orientation trial in Condition D: as C, but one or all stimuli also change in the irrelevant feature dimension (in this case, color).

change magnitude allows for somewhat stronger conclusions on how performance is affected than using highly discriminable stimuli.

3. In most analyses, we will use quantitative process models to support our qualitative conclusions. We use the term *process model* to indicate a model that obtains predictions for psychometric curves by concatenating generic and/or principled assumptions about encoding (storage) and decision (retrieval), instead of postulating a functional form for the psychometric curves without any justification other than goodness of fit. Quantitative process models are useful because they allow us to separate encoding from decision components, because they predict the entire shape of the psychometric curve in each condition rather than only the presence of a difference between conditions, and because they allow for within-subject model comparison.

General experimental design

Throughout this paper, stimuli were colored circular discs (we will say that these stimuli do not have an orientation), oriented white ellipses (we will say for convenience that these stimuli do not have a color), and colored, oriented ellipses (we will say that these stimuli have two feature dimensions); stimulus details are specified below under Experiment 1.

In Experiments 1, 2, and 3, we compared the following four conditions (Figure 1). We used only one condition in a given session, and subjects were instructed on the details of the condition at the start of each session.

Condition A: One feature dimension

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 and red if the response was incorrect. We instructed subjects that their task was to localize the change and that the change could be small or large.

Condition B: Two relevant feature dimensions

In Condition B, each object had both a task-relevant orientation and a task-relevant color. The change occurred in either orientation or color, with equal probability. We instructed subjects that their task was to localize the change in either feature dimension, and that the change could be small or large.

Condition C: One relevant feature dimension and one nonchanging irrelevant feature dimension

Condition C was identical to Condition A except that all objects also had an irrelevant feature dimension. The irrelevant feature dimension did not change between the two arrays. We instructed subjects that their task was to localize the change in the relevant feature dimension, and that the change could be small or large.

Condition D: One relevant feature dimension and one changing irrelevant feature dimension

In Condition D, only one feature dimension was relevant, but an object could change in the irrelevant feature dimension. In one-change Condition D, 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 dimension was relevant and the other was irrelevant; which feature dimension was relevant remained fixed throughout a session for a given subject. The task was to click on the location of the object that had changed in its relevant feature dimension. We instructed subjects that their task was to localize the change in the relevant feature dimension, that the change could be small or large, that one randomly chosen object would change in its irrelevant feature dimensions, and that by chance the same object could change in both dimensions.

All-change Condition D was identical to one-change Condition D, except that on each trial in the second

array, *every* object changed in its irrelevant feature dimension. All changes in the irrelevant feature dimension as well as the change in the relevant feature dimension were independently drawn from a uniform distribution. We instructed subjects that their task was to localize the change in the relevant feature dimension, that the change could be small or large, and that all objects would change in their irrelevant feature dimension. All-change Condition D was designed to magnify any potential effect of the irrelevant feature compared to one-change Condition D.

Models

Qualitative models

Combining the possible answers to Q1 and Q2 results in six models (Figure 2). In Models 1 through 3, orientation and color share memory resource, whereas in Models 4 through 6, orientation and color have independent pools of resource. In Models 1 and 4, an irrelevant feature dimension, 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 dimension 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, we will consider Q3 separately.

Quantitative models: Modeling choices

The models in Figure 2 differ in their qualitative predictions for the four conditions. We will evaluate these predictions using frequentist statistics. To support these analyses, we will also fit quantitative versions of the six models. Quantitative models allow for within-subject model comparisons. Moreover, quantitative *process* models (a) predict entire psychometric curves across conditions, rather than only the presence or absence of an effect of condition on accuracy; (b) allow us to disentangle differences between conditions in encoding precision (of primary interest) from differences between conditions in the decision process (less of interest). The price to pay for these gains is that additional assumptions must be made. Our main assumptions are as follows.

- For our single-feature encoding model, we choose the variable-precision model (Keshvari et al., 2012; Van den Berg et al., 2012; Van den Berg et al., 2014), in which encoding precision is itself a random variable. Although this model is relatively

Model	Memory resource between feature dimensions	Irrelevant feature dimension		Condition A — one feature dimension		Condition B — two relevant feature dimensions			Condition C — one relevant feature + one non-changing irrelevant feature dim.			Condition D — one relevant feature + one changing irrelevant feature dim.			Model ruled out by
				RelF resource	Perf.	RelF₁ resource	RelF₂ resource	Perf.	RelF resource	Effect of IrrF on decision	Perf.	RelF resource	Effect of IrrF on decision	Perf.	
1	Shared	Encoded	Affects decision	⬤	ref.	◗	◗	↓	◗	Small	⇓↓	◗	Large	⇓↓	Exp. 1
2	Shared	Encoded	Ignored	⬤	ref.	◗	◗	↓	◗	None	↓	◗	None	↓	Exp. 1
3	Shared	Not encoded		⬤	ref.	◗	◗	↓	⬤	None	Same	⬤	None	Same	Exp. 2
4	Independent	Encoded	Affects Decision	⬤	ref.	⬤	⬤	Same	⬤	Small	⇓	⬤	Large	↓	
5	Independent	Encoded	Ignored	⬤	ref.	⬤	⬤	Same	⬤	None	Same	⬤	None	Same	
6	Independent	Not encoded		⬤	ref.	⬤	⬤	Same	⬤	None	Same	⬤	None	Same	Shin, Ma (2016)

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

Legend:
- ⬤ A common pool of memory resource (portion: ◗ or ◗)
- ⬤ Resource pool of the first relevant feature dimension
- ⬤ Resource pool of the second relevant feature dimension
- ↓ Performance lowered by resource-sharing
- ⇓↓ Performance lowered by changes in irrelevant feature dimension (small effect: ⇓, large effect: ↓)

RelF: relevant feature dimension
IrrF: irrelevant feature dimension
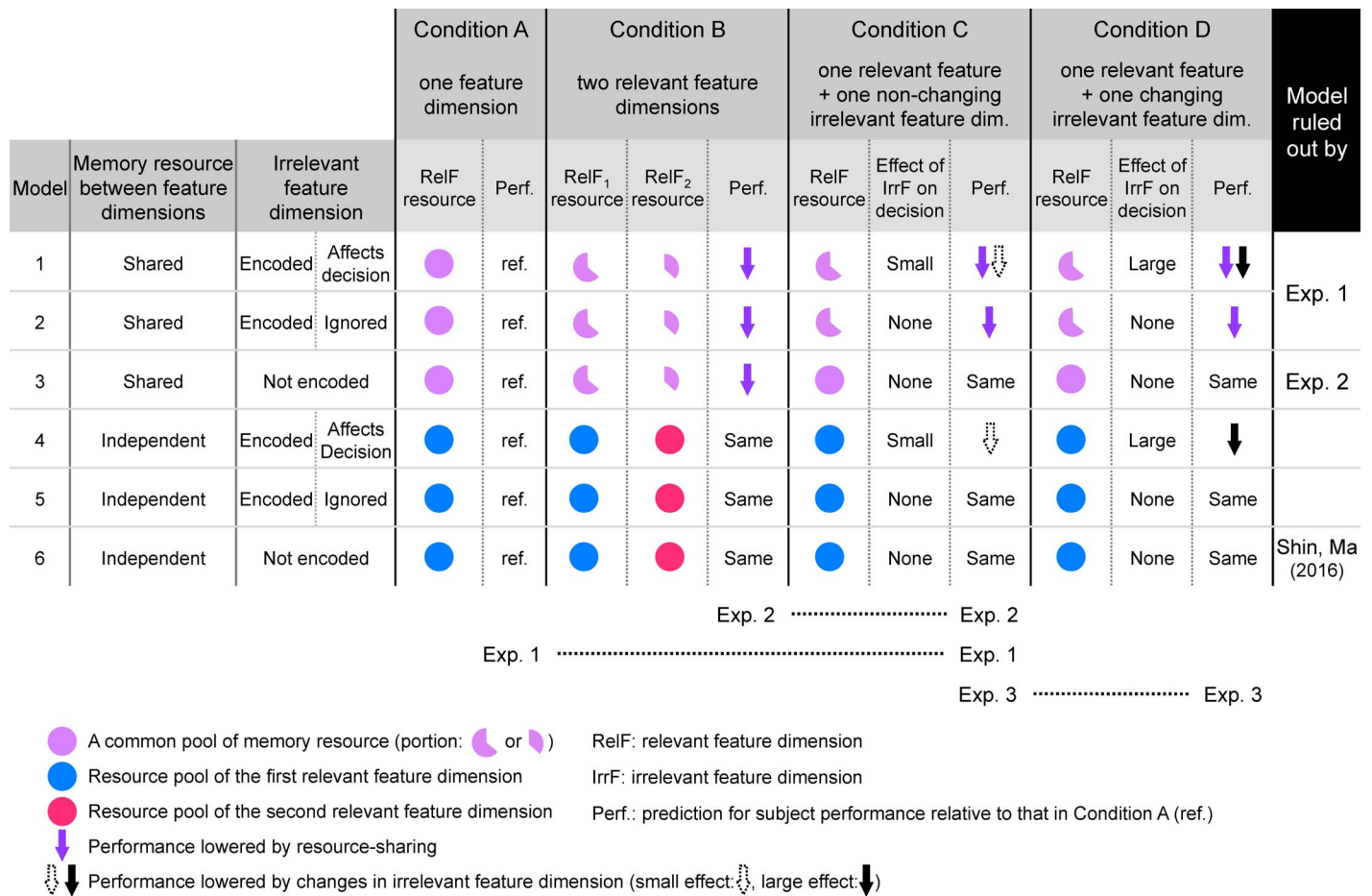Perf.: prediction for subject performance relative to that in Condition A (ref.)

Figure 2. Models of VSTM for Q1 and Q2 and their predictions for performance in different conditions. The models differ in whether memory resource is shared or independent between feature dimensions, and whether an irrelevant feature dimension is encoded and taken into account into decision-making. Discs show how shared (purple) or independent (blue/red) resource is divided across feature dimensions, and the predicted effect on performance (Perf.) relative to performance in Condition A. Below the table, we indicate which conditions are compared in which experiments.

simple and has a good track record (Ma et al., 2014), many other models have been proposed to describe how short-term memories are encoded in a noisy fashion (Bays, 2014; Cappiello & Zhang, 2016; Oberauer & Lin, 2017; Sims, Jacobs, & Knill, 2012; Van den Berg et al., 2014; Zhang & Luck, 2008). The present study cannot rule those models out, and at least qualitatively, the differences between conditions predicted by the variable-precision model can also be accounted for by the other models.

- For the observer's decision, we choose a Bayes-optimal (ideal) observer. This decision model described human data best in a similar change detection task (Keshvari et al., 2012), but other, suboptimal rules, can certainly not be ruled out.

These assumptions are generic, and the second one could be called principled. Moreover, we will see that models equipped with these assumptions can yield visually good fits to the psychometric curves. Never-

theless, it is important to keep in mind that the present work cannot and is not aimed at ruling out any alternatives for the encoding and decision stage beyond the aspects specified in Figure 2.

Quantitative models: One feature dimension (Condition A)

We first focus on objects that have a single feature dimension. We model the observer's change localization decisions using a Bayes-optimal model. We derived the Bayesian decision model for change localization previously (Van den Berg et al., 2012) but we briefly summarize the logic here. 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). After describing these three steps for single-feature objects, we will describe the modifications necessary for two-feature objects.

Step 1: Encoding

The stimulus arrays before and after the change both contain N objects. We denote the vector of stimuli in the first array by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$. Every θ_i is independently drawn from a uniform distribution. In going from the first to the second stimulus array, exactly one object (drawn with equal probabilities) changes its value; we denote the location where this happens by L. The magnitude of the change, denoted by Δ, is drawn from a uniform distribution. The vector of stimuli in the second array is denoted by $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_N)$. Since the change occurs in one object, $\boldsymbol{\varphi}$ and $\boldsymbol{\theta}$ are identical except for the Lth entry, where $\varphi_L = \theta_L + \Delta$.

When the stimulus in the first array is θ_i, the observer has a noisy measurement (memory) of this stimulus, which we denote by x_i. Similarly, we denote the noisy measurement of φ_i by y_i. We denote the measurement vectors by $\mathbf{x} = (x_1, \ldots, x_N)$, and $\mathbf{y} = (y_1, \ldots, y_N)$. We assume that the noise corrupting the measurements is independent across arrays and locations. Since orientation and color are circular variables in our experimental design, we assume that x_i and y_i follow Von Mises distributions (with orientation rescaled to have the range [0,2π]):

$$p(x_i|\theta_i) = \frac{1}{2\pi I_0(\kappa_{x,i})} e^{\kappa_{x,i}\cos(x_i-\theta_i)},$$

$$p(y_i|\varphi_i) = \frac{1}{2\pi I_0(\kappa_{y,i})} e^{\kappa_{y,i}\cos(y_i-\varphi_i)}. \quad (1)$$

Here, I_0 is the modified Bessel function of the first kind of order zero (Abramowitz & Stegun, 1972), and $\kappa_{x,i}$ and $\kappa_{y,i}$ are called concentration parameters, which we will assume to be stochastic themselves (see below). Both the Von Mises and the independence assumptions are simplifications, but they fit the present data well. Moreover, most of our qualitative conclusions are supported by model-free statistics and are unlikely to be sensitive to changes in model assumptions.

From resource to encoding precision

We 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.

Per the variable-precision model (Fougnie et al., 2012; Van den Berg et al., 2012; Van den Berg et al., 2014), we allow the top-down resource factor at the ith location in a stimulus array, denoted by $J'_{\text{array},i}$, to be variable across arrays (first and second), locations (objects), and trials.

Specifically, we 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, we 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 ith location in a given array:

$$J_{\text{array},i} = \alpha J'_{\text{array},i}. \quad (2)$$

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

$$J_{\text{array},i} = \kappa_{\text{array},i} \frac{I_1(\kappa_{\text{array},i})}{I_0(\kappa_{\text{array},i})} \quad (3)$$

where I_1 is the modified Bessel function of the first kind of order one (Abramowitz & Stegun, 1972). This relationship, which is nearly linear, follows from the interpretation of precision as Fisher information (Van den Berg et al., 2012; Van den Berg et al., 2014).

Step 2: Inference

We denote by d_L the likelihood ratio, based on x_L and y_L, that a change occurred at the Lth location, disregarding all other locations. It turns out that an accuracy-maximizing observer will report the location for which d_L is highest (see Appendix A).

Step 3: Response probabilities

In our model, each trial in a single-feature condition is characterized completely by its change magnitude Δ; the values of stimuli $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ 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, $\boldsymbol{\omega}$, and a given change magnitude Δ, we calculated the probability of a correct response, denoted by $p(\text{correct}|\Delta, \boldsymbol{\omega})$, through Monte Carlo simulation. This entailed the following. We generated 1,280 samples of $J_{x,i}$ and $J_{y,i}$ for the first and second stimulus array according to the process described under Step 1. We used these values to compute concentration parameters $\kappa_{x,i}$ and $\kappa_{y,i}$ according to Equation 3, and we 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 1,280 combinations of κ_x, κ_y, $\mathbf{x}$, and $\mathbf{y}$ thus drawn, we evaluated the decision rule. Tallying correct responses across these draws yielded an estimate of the probability of a correct localization response for the given parameter combination and the given change magnitude. We used these probabilities for model fitting.

Quantitative models: Two feature dimensions (Conditions B, C, and D)

We now describe the modifications necessary for objects with two features, say orientation and color.

Step 1: Encoding

The stimuli on a given trial are represented by a quadruplet (θ_{ori}, θ_{col}, φ_{ori}, φ_{col}). In Condition B, both features are relevant, and a change occurs in one feature of one object. In Conditions C and D, there is a relevant and an irrelevant feature. In Condition C, the irrelevant feature, denoted by a subscript "irr", never changes, and thus, $\varphi_{\text{irr}} = \theta_{\text{irr}}$. In Condition D, the irrelevant feature might change. In the one-change condition, an irrelevant change of magnitude Δ_{irr} is introduced at the Lth 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: $\boldsymbol{\varphi}_{\text{irr}} = \boldsymbol{\theta}_{\text{irr}} + \boldsymbol{\Delta}_{\text{irr}}$, where $\boldsymbol{\Delta}_{\text{irr}} = (\Delta_{1,\text{irr}} \ldots, \Delta_{N,\text{irr}})$, 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. We assume that the noise corrupting the orientation and color measurements is independent across arrays, locations, and features (Fougnie & Alvarez, 2011).

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

In the shared-resource models, orientation and color share a common pool of resource. Specifically, we assume that the sum of the mean resource of orientation and that of color is equal to a fixed value $\bar{J}'$. Thus, we let the mean orientation resource be $\rho \bar{J}'$, and that of color $(1 - \rho)\bar{J}'$, where $0 < \rho < 1$. We assume

independent variability for orientation and color:

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

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

where τ'_{ori} and τ'_{col} are the scale parameters for orientation and color, respectively. We 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

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

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

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, we 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: $\bar{J}_{\text{ori}}$, $\bar{J}_{\text{col}}$, τ_{ori}, τ_{col}, and ρ.

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 *relevant* orientation and color follow:

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

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

In Models 4 and 5, irrelevant features have a pool of resource, whereas in Model 6, they do not. In Models 4 and 5, we do not assume that the amount of resource for the irrelevant feature is the same as when that feature is relevant. Instead, we allow it to be a proportion. In Model 5, those proportions do not need to be explicitly modeled, since the irrelevant feature is not taken into account. In Model 4, however, we do need to model these proportions separately for orientation and color, giving rise to parameters ρ_{ori} and ρ_{col}. In Model 4, mean precision for *irrelevant* orientation is then $\rho_{\text{ori}}\bar{J}_{\text{ori}}$, and mean precision for *irrelevant* color is $\rho_{\text{col}}\bar{J}_{\text{col}}$; we assume that τ_{ori} and τ_{col} are identical to when the features are relevant.

All independent-resource models have the four parameters $\bar{J}_{\text{ori}}$, $\bar{J}_{\text{col}}$, τ_{ori}, and τ_{col}, and Model 4 has two more, ρ_{ori} and ρ_{col}.

Step 2: Inference

In Condition B (two relevant feature dimensions), the decision variable turns out to be simply the average of the likelihood ratios for the individual features (see Appendix A). The observer reports the location for which this average is highest. In Conditions C and D (one relevant, one irrelevant feature dimension), if the irrelevant feature dimension is ignored during decision-making, the decision rule is as in Condition A, whereas if it is taken into account, we assume that the decision rule is the same as in Condition B.

Step 3: Response probabilities

For each model for two-feature objects, we 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 (e.g., in Condition C, one of the vectors is the zero vector, whereas in Condition D, neither is). In each experimental condition, we estimated the model observer's probability of a correct response, $p(\text{correct}|\Delta_{ori}, \Delta_{col}, \omega)$, through Monte Carlo simulations with 1,280 samples of $\kappa_{ori,x}$, $\kappa_{ori,y}$, x_{ori}, y_{ori}, $\kappa_{col,x}$, $\kappa_{col,y}$, x_{col}, and y_{ori}.

Model fitting

For here and elsewhere, we used maximum-likelihood estimation for model fitting. The likelihood of a parameter combination ω is $L(\omega) = p(\text{data}|\omega)$. We maximized this function over ω, which is equivalent to maximizing its logarithm. Assuming that all trials are conditionally independent, the log-likelihood function for a given condition is

$$\log L(\omega) = \log p(\text{data}|\omega)$$
$$= \log \prod_{t=1}^{N_{trials}} p(\text{correctness}_t|\Delta_{ori,t}, \Delta_{col,t}, \omega)$$
$$= \sum_{t=1}^{N_{trials}} \log p(\text{correctness}_t|\Delta_{ori,t}, \Delta_{col,t}, \omega),$$

where N_{trials} is the number of trials in the condition (in Condition A, only one of either $\Delta_{ori,t}$ and $\Delta_{col,t}$ exists). We fitted all conditions within an experiment simultaneously, and thus the log-likelihood gets summed across all conditions. For maximizing the resulting total log likelihood, we used the genetic algorithm ("ga") in the Global Optimization Toolbox in MATLAB, with 64 individuals and 64 generations.

Overview of experiments

In the Introduction, we formulated three central questions, Q1 through Q3. Models 1 through 6 capture possible answers to questions Q1 and Q2. In Experiment 1, we test and rule out Models 1 and 2. In Experiment 2, we test and rule out Model 3. To rule out Model 6, we refer to a published study using a delayed-estimation task. In Experiment 3, we attempt unsuccessfully to distinguish Models 4 and 5. Finally, in Experiment 4, we answer question Q3.

Experiment 1: Ruling out Models 1 and 2

Experiment 1 tests Models 1 and 2, in which the irrelevant feature receives a portion of a shared resource. In this experiment, subjects were tested on Conditions A (one feature) and C (a relevant and a nonchanging irrelevant feature) using change localization tasks (Figure 1). Models 1 and 2 both postulate that memory resource is shared between the features in Condition C. Therefore, they 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 (Figure 2).

Methods

Stimuli and subjects: Stimuli were displayed on a 19″ LCD monitor at a viewing distance of approximately 60 cm. Stimuli were presented on a mid-level gray background of luminance 33.1 cd/m². Stimuli were equally spaced along an imaginary circle of radius 7 degrees of visual angle (deg) around fixation (calculated assuming a viewing distance of 60 cm), at angles $[45 + (I - 1)/N \times 360]°$, where $i = 1, \ldots, N$, and $N = 4$. All experiments were programmed using Psychophysics Toolbox in MATLAB (Brainard, 1997; Pelli, 1981).

An orientation-only stimulus was a white ellipse of luminance 95.7 cd/m² 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/m², with center $(a^*, b^*) = (12, 13)$ and radius 60. A two-feature stimulus was a colored ellipse.

Eight subjects (age range 23–30 years), including author H. S., participated. Besides the author, all subjects were naïve to the goal of the experiment. All subjects gave informed consent. The experimental protocol adhered to the Declaration of Helsinki and
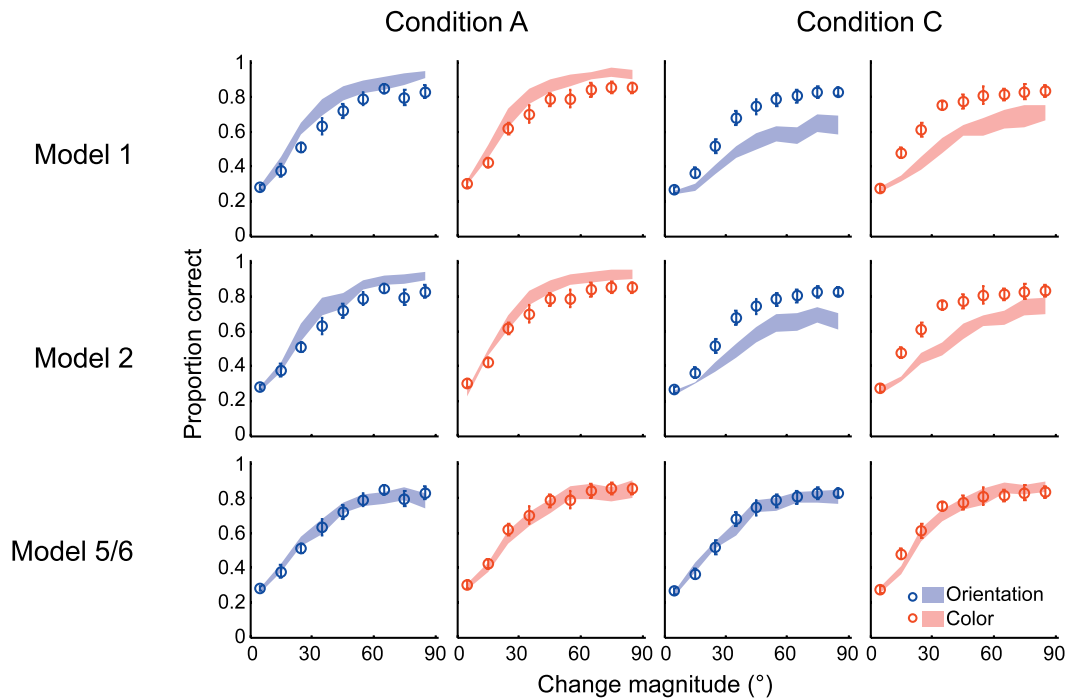
Figure 3. Experiment 1. Proportion correct as a function of change magnitude (eight subjects), with model fits (shaded areas: $M \pm$ SEM). Performance was indistinguishable between Conditions A and C, in accordance with Model 5/6 but not with Models 1 and 2.

was approved by the Institutional Review Board of Baylor College of Medicine.

Procedure: Experiment 1 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). The order of the sessions was random for each subject. Each session consisted of four blocks of 150 trials. Hence, each subject completed $4 \times 4 \times 150 = 2,400$ trials in total. At the beginning of each session, subjects completed 10 practice trials. Each session lasted about 45 min. Set size was 4.

Results

Psychometric curves are shown in Figure 3. For orientation, a two-variable logistic regression of accuracy against change magnitude (a continuous variable) and condition (binary) revealed a significant effect of change magnitude, $\beta = 0.035 \pm 0.002$; t test on β values: $t(7) = 6.89$, $p < 0.001$, but no significant effect of condition, $\beta = -0.018 \pm 0.029$, $t(7) = -0.21$, $p = 0.84$. For color, we similarly found a significant effect of change magnitude, $\beta = 0.035 \pm 0.002$, $t(7) = 7.23$, $p < 0.001$, and no significant effect of condition, $\beta = -0.128 \pm 0.028$, $t(7) = -1.53$, $p = 0.17$. Since Models 1 and 2 predict a significant effect of condition, we find no evidence for Models 1 or 2.

Since a frequentist test cannot prove a null hypothesis, we also performed formal model comparison. We compared Models 1 and 2 against Model 5/6 (Models 5 and 6 make the same prediction for this experiment). For every individual subject, the Akaike information criterion (AIC) of Models 1 and 2 are higher than that of Models 5/6, on average by 210 ± 13 and 154 ± 11 ($M \pm SEM$), respectively. This confirms that we can reject Models 1 and 2.

Experiment 2: Ruling out Model 3

So far, we have ruled out Models 1 and 2. Next, we test Model 3, in which orientation and color share a memory resource but the irrelevant feature is not encoded. To this end, we compare Conditions B (two relevant features, one of which changes) and C (a relevant and a nonchanging irrelevant feature; Figure 1).

Methods

Stimuli and subjects: Stimuli were identical to those in Experiment 1. Eight subjects (including one author) participated and the age range was between 23 and 30 years.

Procedure: Experiment 2 consisted of four sessions, run on different days: two sessions of Condition B (statistically identical to each other) and two sessions of
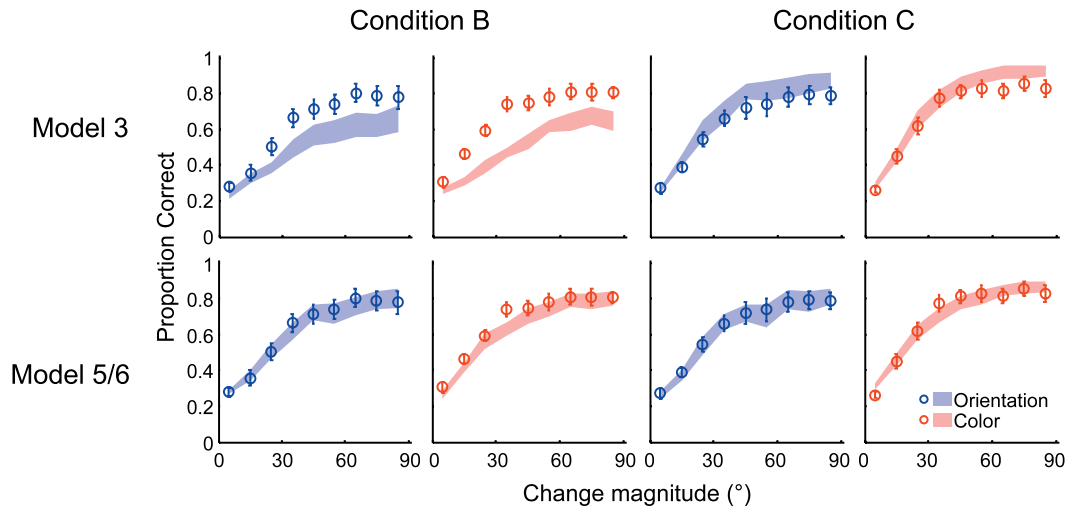
Figure 4. Experiment 2. Proportion correct as a function of change magnitude (eight subjects), with model fits (shaded areas: $M \pm$ SEM). Performance was indistinguishable between Conditions B and C, in accordance with Model 5/6 but not with Model 3.

Condition C (of which one was a relevant-orientation session and one a relevant-color session). Set size was 4. We conducted the same experiment at a higher set size ($N = 8$). Otherwise, Experiment 2 was the same as Experiment 1.

Results

Psychometric curves are shown in Figure 4. For orientation, a two-variable logistic regression of accuracy against change magnitude (a continuous variable) and condition (binary) revealed a significant effect of change magnitude, $\beta = 0.039 \pm 0.001$; t test on β values: $t(7) = 9.01$, $p < 0.001$, but no significant effect of condition, $\beta = 0.025 \pm 0.029$, $t(7) = 0.28$, $p = 0.79$. For color, we similarly found a significant effect of change magnitude, $\beta = 0.037 \pm 0.002$, $t(7) = 8.00$, $p < 0.001$, and no significant effect of condition, $\beta = -0.022 \pm 0.032$, $t(7) = -0.23$, $p = 0.83$.

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, we find no evidence for Model 3. In model comparison, we found that for every individual subject, the AIC of Model 3 is higher than that of Models 5/6, on average by 161 ± 32. We confirmed these findings in a separate experiment (5 subjects) that was identical except that set size was 8 (AIC difference: 81 ± 19; Appendix B). Therefore, we can reject Model 3.

Answer to Q1

Taking the results of Experiments 1 and 2 together, we can rule out that, for our stimuli, memory resource is shared between orientation and color. These findings

are consistent with classic change detection studies (Luck & Vogel, 1997; Olson & Jiang, 2002; Vogel et al., 2001). Another study that worked within the noisy-memory framework (Fougnie, Asplund, & Marois, 2010) obtained somewhat mixed results: 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 also inconclusive: A cost for remembering two features over one was observed when the change was small but not when it was large (Fougnie et al., 2010). By contrast, our results are consistent between features and across change magnitudes.

Ruling out Model 6

We are now left with the three independent-resource models (Models 4, 5, and 6). Of these, Model 6 distinguishes itself by the assumption that the irrelevant feature is not encoded. Unfortunately, in change detection and change localization, it is fundamentally impossible to distinguish between "not encoded" and "encoded but not used." This same problem applies to classic change detection studies that claimed that the irrelevant feature is not encoded (Luria & Vogel, 2011; Vogel et al., 2001). Therefore, we have to probe the encoding using a more direct paradigm.

We did this in recent work (Shin & Ma, 2016) by using a delayed-estimation task (Blake, Cepeda, & Hiris, 1997; Nilsson & Nelson, 1981; Wilken & Ma, 2004) on the Amazon Mechanical Turk online data
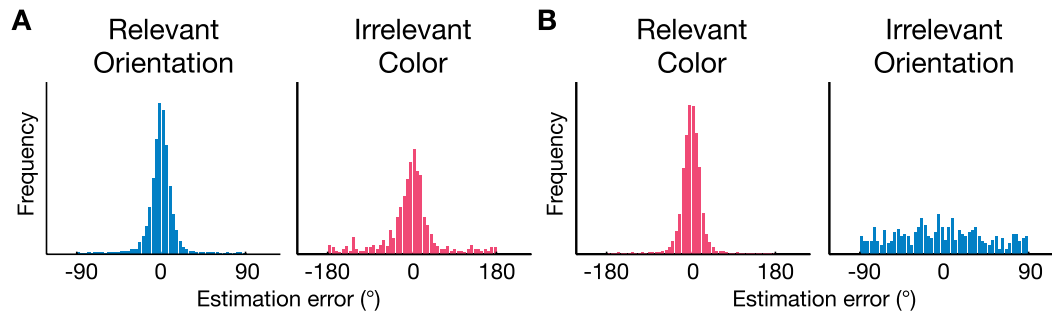
Figure 5. Results from Shin and Ma (2016). The task was delayed estimation. In each experiment, 600 subjects each performed 30 relevant-feature trials and one irrelevant-feature trial. Thus, the relevant-feature histograms are based on 18,000 trials and the irrelevant-feature histograms on 600 trials. (A) Orientation was relevant and color irrelevant. Both distributions are significantly different from uniform. (B) Color was relevant and orientation was irrelevant. Both distributions are significantly different from uniform.

collection platform. On each trial, the stimulus was a colored ellipse, with orientation and color independently drawn from their respective uniform distributions. After a delay, subjects ($n = 600$) 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 30 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. We found that for both orientation and color, the error distribution on irrelevant-feature trials was not uniform (Figure 5). This indicates that subjects do encode the irrelevant feature and rules out Model 6. However, the quality of encoding of the irrelevant feature was rather poor: The inverse of the square of the circular standard deviation when orientation was irrelevant was only 3.8% of the same quantity when orientation was relevant, and for color, that ratio was 20.4%.

By contrast, in a recent study, the proportion of participants responding correctly on a surprise trial that required identification of a previously irrelevant color was not significantly different from chance (Chen & Wyble, 2015). Potential reasons why weak encoding was not detected by this study are that the number of subjects was only 22 (we had 600 in our study), and they used a four-choice paradigm rather than the more informative continuous-estimation paradigm.

Woodman and Vogel (2008) asked whether irrelevant information was represented in VSTM by comparing the amplitude of the component of the event-related potential known as the contralateral delay activity between conditions in which only color, only orientation, or both were relevant. They found an effect of condition, suggesting that relevance affected the strength of encoding. However, this result does not rule out that irrelevant features are weakly encoded.

Our finding can be contrasted with a study that found that the color of an irrelevant stimulus was reported incorrectly by 18% of participants (Eitam, Yeshurun, & Hassan, 2013). Despite the title of their paper ("Blinded by irrelevance"), this means that the irrelevant color was reported correctly by the vast majority of participants, suggesting strong encoding. However, presentation time was much longer (500 ms) and relevance was introduced by a verbal manipulation ("concentrate on the inner circle"), which might be less effective than reinforced task demands; also, both features were colors.

Experiment 3: Comparing Models 4 and 5

Two models remain: Models 4 and 5. In both models, the irrelevant feature is encoded (although potentially with only a fraction of the amount of resource allocated when that feature is relevant), but in Model 4, it is taken into account in decision-making, whereas in Model 5, it is ignored during decision-making. We conducted Experiment 3 to distinguish between these possibilities. In Experiment 3, we followed the concept of Hyun et al. (2009) and introduced changes in the irrelevant feature. We tested subjects on Condition C (a relevant and a nonchanging irrelevant feature) and Condition D, in which the value of the irrelevant feature of an object changed.

Methods

Stimuli and subjects: Stimuli were identical to those of Experiment 1. Five subjects (including one author) participated, and the age range was between 23 and 30 years.
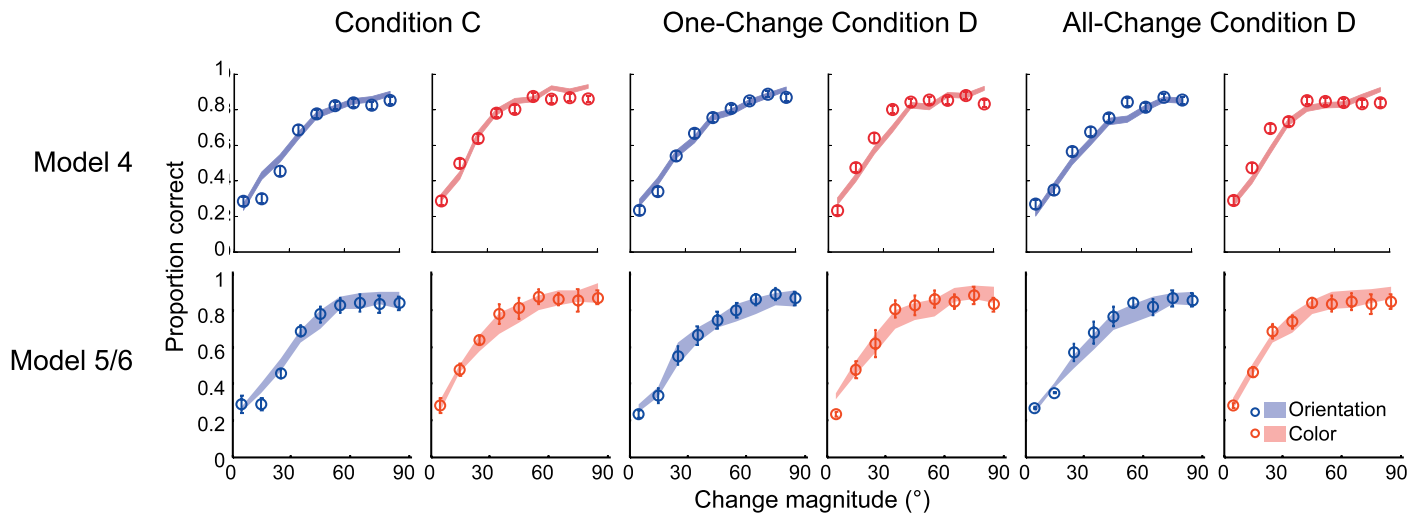
Figure 6. Experiment 3. Proportion correct as a function of change magnitude (five subjects), with model fits (shaded areas: $M \pm$ SEM). Performance was indistinguishable between Conditions C and D, but both Model 4 and Model 5 fit well.

Procedure: Experiment 3 consisted of six sessions, all including a relevant-orientation session and a relevant color session: two sessions of no-change Condition C, two sessions of one-change Condition D, and two sessions of all-change Condition D. Otherwise, Experiment 3 was the same as Experiment 1.

Results

Psychometric curves are shown in Figure 6. Since we have three conditions, we perform logistic regression of accuracy against change magnitude, a dummy variable for Condition C, and a dummy variable for one-change Condition D. For orientation, this revealed a significant effect of change magnitude, $\beta = 0.044 \pm 0.002$; t test on β values: $t(4) = 7.58$, $p = 0.002$; no significant difference between Condition C and the other conditions, $\beta = -0.066 \pm 0.044$, $t(4) = -0.60$, $p = 0.58$); and no significant difference between one-change Condition D and the other conditions, $\beta = 0.034 \pm 0.036$, $t(4) = 0.37$, $p = 0.73$. For color, we similarly found a significant effect of change magnitude, $\beta = 0.041 \pm 0.003$, $t(4) = 6.30$, $p = 0.003$; no significant difference between Condition C and the other conditions, $\beta = -0.050 \pm 0.056$, $t(4) = -0.03$, $p = 0.98$; and no significant difference between one-change Condition D and the other conditions, $\beta = -0.042 \pm 0.040$, $t(4) = -0.42$, $p = 0.70$. The absence of an effect of an (either changing or nonchanging) irrelevant feature on performance is consistent with most change detection studies that did not vary change magnitude (Luria & Vogel, 2011; Shen et al., 2013; Vogel et al., 2001; Xu, 2010). It is inconsistent with (Hyun et al., 2009), and we do not know why.

Models 4 and 5 agree that the irrelevant feature is encoded, but differ in whether it is taken into account

in decision-making. Model 5 predicts that performance in all three conditions is identical, while Model 4 predicts that any change in the irrelevant feature decreases performance, and the more objects change in their irrelevant feature, the greater the decrease. Therefore, Model 4 predicts that performance is lower in one-change Condition D than in Condition C, and even lower in all-change Condition D. Thus, the logistic regression does not provide any evidence for Model 4.

However, it is possible that the irrelevant feature is encoded with low precision, as suggested by the Amazon Turk experiment; the logistic regression might not be able to detect an effect of condition on performance. Therefore, we conduct formal model comparison. We find that the AIC of Model 4 is higher than of Model 5 by 37 ± 23, reflecting substantial intersubject differences. In Model 4, we estimate the parameter ρ_{ori}, which represents the precision with which orientation is encoded when irrelevant as a proportion of the precision with which it is encoded when relevant, as $12.9\% \pm 3.3\%$ ($t = 3.88$, $p = 0.02$). We estimated the analogous parameter for color, ρ_{col}, as $11.6\% \pm 8.4\%$ ($t = 1.38$, $p = 0.24$). Therefore, we cannot state conclusively whether the irrelevant feature is taken into account, but if it is, it is encoded with comparatively low precision and therefore has little effect on the decision.

Answer to Q2

Experiments 1 through 3 narrow the possible models down to Models 4 and 5: The irrelevant feature dimension is encoded but either ignored, or encoded with such low precision that it barely affects performance.
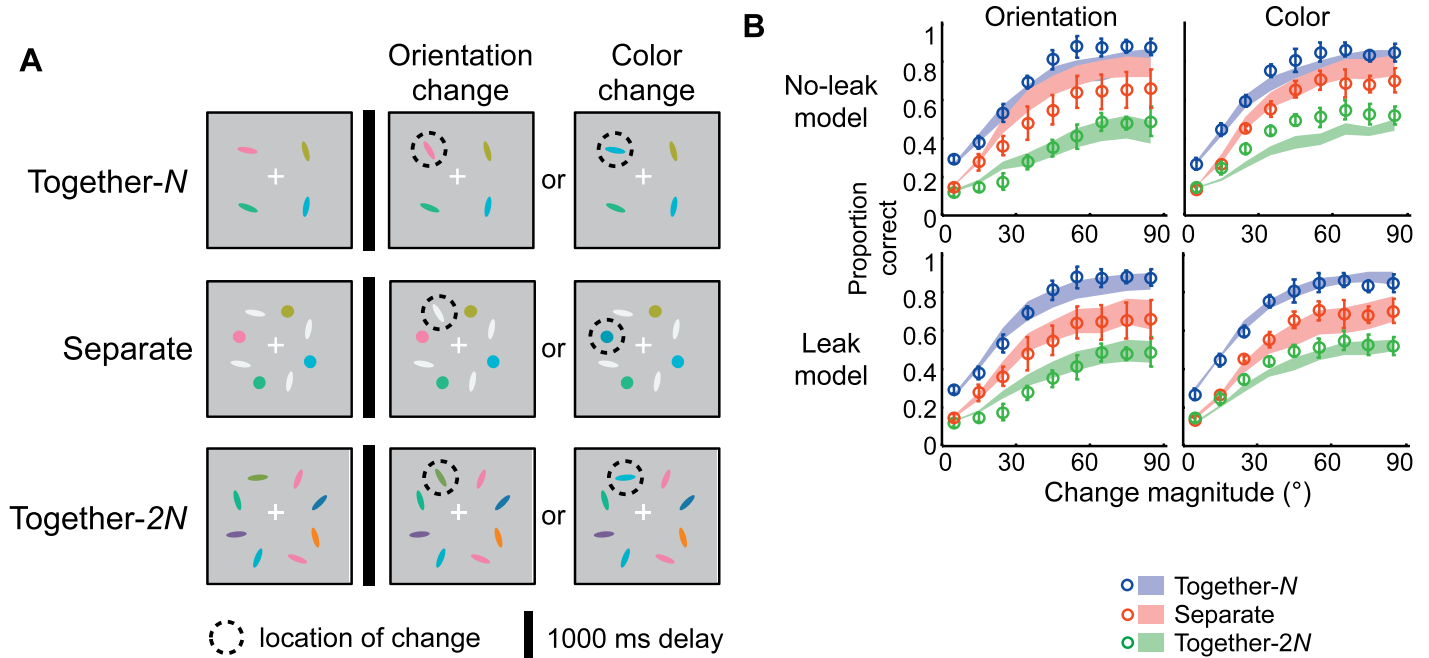
Figure 7. Experiment 4. (A) Trial procedure. A change occurred either in orientation or color. Subjects clicked on the location where the change occurred. The three conditions differed in set size (4 or 8) and the number of features per object (1 or 2). (B) Psychometric curves and model predictions. The no-leak model (top) fits better than the leak model (bottom).

Experiment 4

The remaining question, Q3, is whether dividing the features of N two-feature objects over $2N$ single-feature objects decreases the amount of resource in a given feature dimension that is available to encode the N feature values in that dimension. A possible mechanism for such a decrease would be that some amount of resource "leaks" to the N "distractor" objects that do not have that feature dimension but are task-relevant because of their other feature dimension. This question is orthogonal to Q1 and Q2, and independent of the models considered so far. We examined Q3 in Experiment 4, using three conditions that parallel those in Olson and Jiang (2002).

Methods

Stimuli and subjects: Stimuli are identical to those of Experiment 1. Five subjects (including one author) participated, with an age range of 26–30 years. *Conditions and procedure*: Experiment 4 consisted of two together-N, two separate, and two together-$2N$ sessions (Figure 7A). 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 $6 \times 4 \times 150 = 3{,}600$ trials in total. 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 before). 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 eight objects. The together-$2N$ condition was identical to the together-N condition except that set size was 8.

Models

We consider two models, both of them variants of Model 5 (Figure 7B); since Models 4 and 5 are in practice indistinguishable, we expect the same results under Model 4. In the *no-leak model*, the entire amount of resource in a given feature dimension is distributed to relevant locations (i.e. to objects that have that feature dimension). In the *leak model*, only a portion of the resource is distributed to relevant locations.

In both models, we use the assumptions about encoding precision and decision-making of Model 5: Each feature has its own resource, and irrelevant features are encoded but ignored in decision-making. 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.

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 eight decision variables: four from orientation and four from color. This is comparable to a change localization task with eight one-feature objects, although here there are four colored 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: $\bar{J}_{\mathrm{ori}}, \bar{J}_{\mathrm{col}}, \tau_{\mathrm{ori}}, \tau_{\mathrm{col}}$.

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: $\bar{J}_{\mathrm{ori}}, \bar{J}_{\mathrm{col}}, \tau_{\mathrm{ori}}, \tau_{\mathrm{col}}, r_{\mathrm{ori}}$, and r_{col}.

Results

Psychometric curves are shown in Figure 7B. 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. 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 compared to the leak model with independent leak per

feature (AIC difference: 1.9 ± 1.6) and this leak rate is estimated at $31 \pm 4\%$.

Answer to Q3

We found that resource for a given feature dimension is lower in the presence of task-relevant distractors, stimuli that do not have that feature dimension. One possible mechanism is that feature resource leaks to those distractors. Both the finding and the possible mechanism are consistent with a recent delayed-estimation study (Fougnie et al., 2010), as well as with older change detection studies (Lee & Chun, 2001; Olson & Jiang, 2002); however, of these, only the latter included 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 might "leak," it is still far from being equally allocated to all objects regardless of relevance.

Discussion

The binary question, "Is VSTM object-based or feature-based?" has always been multifaceted. Here, we performed a series of experiments to disentangle these facets within a noise- or resource-based framework. Our main conclusions are: (Q1) Color and orientation have independent pools of resource; (Q2) An irrelevant feature is encoded, but with much lower precision than when the same feature is relevant, and it is not clear whether it is taken into account during decision-making; (Q3) Resource allocated to objects that have a given feature dimension is lower in the presence of task-relevant objects that do not have that feature dimension, suggesting that some amount of resource leaks to the latter objects.

Our answer to Q1 is puzzling in light of recent studies that found that change detection accuracy decreased with the number of features (Hardman & Cowan, 2015; Oberauer & Eichenberger, 2013). Several potential explanations for these discrepancies come to mind: (a) The number of feature dimensions: we used only two feature dimensions, whereas those studies used up to six dimensions. It is possible that resource pools are independent only when the number of features is low. (b) The identities of the features: aside from orientation and color, Oberauer and Eichenberger (2013) also used thickness, frequency, shape, and size, and Hardman and Cowan (2015) also used length, and presence of a gap; it is possible that certain feature pairs (for example pairs of spatial features) share resource, even if color and orientation
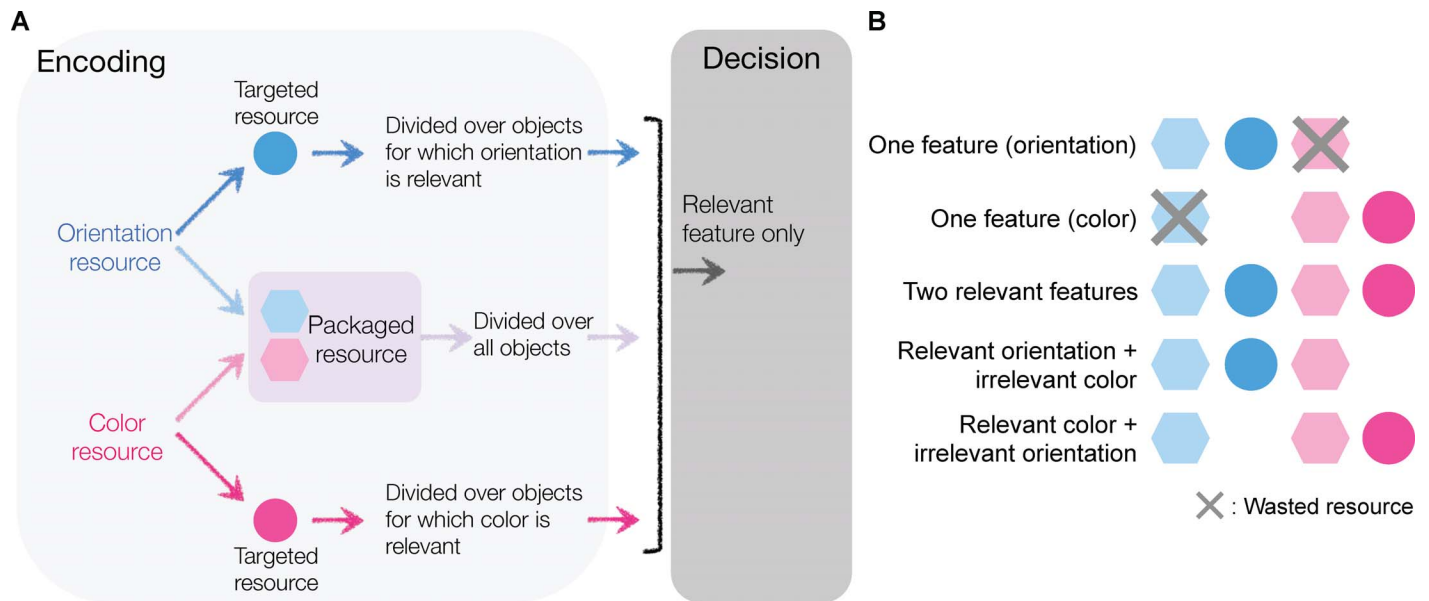
Figure 8. Orientation-color VSTM in a noisy-memory framework. (A) Feature resource consists of "packaged" resource, which gets allocated to all objects, and "targeted" resource, which only gets allocated to objects for which that feature is relevant. (B) Application to the conditions from Figure 1.

do not. (c) In a noisy-memory framework, the decision stage might be a confound. We draw a parallel with visual search, in which an additional memory feature would correspond to an additional distractor, and independent resource pools would correspond to precision per stimulus being independent of the number of stimuli. In noise-based visual search models, adding distractors reduces accuracy on whole-array detection of a single target *even if encoding precision per stimulus is independent of the number of stimuli* (Eckstein, 1998; Nolte & Jaarsma, 1967; Verghese, 2001). The reason is that the noise from the distractors increasingly "drowns out" the target signal. In the present study, we avoided this pitfall by using appropriately different decision rules in Condition A (single feature) and Condition B (two relevant features) in the models. Further work is needed to distinguish between these possibilities.

The question arises whether an overarching framework for multi-feature VSTM can unify our findings and those of others. Brady, Konkle, and Alvarez (2011) have suggested that the unit of VSTM is a "hierarchically structured feature bundle" (p. 7). 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." Here, we propose a conceptually related but different framework (Figure 8A). Each feature has an independent pool of VSTM resource. Some of the resource for a given feature gets placed into multifeature 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 (we use the term *package* for resources, in contrast to Brady et al.'s [2011] term *bundle* for features). Within the multifeature package, resource remains feature-specific, and if the object does not have one of the features, the corresponding resource from the package ends up being wasted. Finally, the brain utilizes a smart decoder, in the sense that the signal arising from resource being allocated to an irrelevant feature dimension can be ignored during decision-making.

This framework can account for all our findings as well as some others (Fougnie et al., 2010; Marshall & Bays, 2013; Figure 8B). 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. Our recent study (Shin & Ma, 2016) suggests that 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 targeted resource and performance is lower. In Experiment 4, 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 would be consistent with the results of (Fougnie et al., 2010; Lee & Chun, 2001; Olson & Jiang, 2002) and corresponds to an object benefit (Jiang et al., 2000) for the together-*N* condition. Moreover, precision in the separate condition is higher than in the together-2*N* 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 Marshall and Bays (2013) and with classic change detection results (Olson & Jiang, 2002). However, it is inconsistent with one comparison in (Marshall & Bays, 2013) that used a condition in which two sequentially presented displays differed by which feature was relevant (e.g., 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.

An aspect of VSTM that we have not addressed is binding errors (Bays, 2016; Bays, Catalao, & Husain, 2009; Oberauer & Lin, 2017; Van den Berg et al., 2014; Wheeler & Treisman, 2002): a feature might be correctly remembered but at the wrong location. These errors could already occur in single-feature VSTM. They can be thought of as arising from the noisy representation of location, with confusions more likely when stimuli are closer together (Bays, 2016). In most of the work presented here, set size was 4, with stimulus locations being far apart and fixed across the experiment. These design features are likely to minimize binding errors. The exceptions were the separate and together-$2N$ conditions in Experiment 4, in which set size was 8. We leave the study of potential binding errors in these conditions for future work.

Taken together, we have shown that applying the framework of noisy memories to questions regarding multiple-feature VSTM brings along new techniques (full psychometric curves over change magnitude for stronger model tests; model-based separation of encoding and decision stage), confirms some long-standing notions, offers a new perspective on others, and introduces new questions.

Keywords: visual working memory, visual short-term memory, objects, features, computational modeling, change detection

References

Abramowitz, M., & Stegun, I. A. (Eds.). (1972). *Handbook of mathematical functions.* New York: Dover Publications.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, *34*(10), 3632–3645.

Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports*, *6*: 19203.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10):7, 1–11, doi:10.1167/9.10.7. [PubMed] [Article]

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854.

Blake, R., Cepeda, N. J., & Hiris, E. (1997). Memory for visual motion. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 353–369.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5):4, 1–34, doi:10.1167/11.5.4. [PubMed] [Article]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Cappiello, M., & Zhang, W. (2016). A dual-trace model for visual sensory memory. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(11), 1903–1922.

Chen, H., & Wyble, B. (2015). The location but not the attributes of visual cues are automatically encoded into working memory. *Vision Research*, *107*, 76–85.

Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science*, *9*(2), 111–118.

Eitam, B., Yeshurun, Y., & Hassan, K. (2013). Blinded by irrelevance: Pure irrelevance induced "blindness." *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 611–615.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(12):3, 1–12, doi:10.1167.11.12.3. [PubMed] [Article]

Fougnie, D., Asplund, C. L., & Marois, R. (2010).

What are the units of storage in visual working memory? *Journal of Vision*, *10*(12):27, 1–11, doi:10.1167/10.12.27. [PubMed] [Article]

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

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 325–347.

Hyun, J., Woodman, G. F., Vogel, E. K., Hollingworth, A., & Luck, S. J. (2009). The comparison of visual working memory representations with perceptual inputs. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1140–1160.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 683–702.

Keshvari, S., Van den Berg, R., & Ma, W. J. (2012). Probabilistic computation in human perception under variability in encoding precision. *PLoS ONE*, *7*(6), e40216.

Keshvari, S., Van den Berg, R., & Ma, W. J. (2013). No evidence for an item limit in change detection. *PLoS Computational Biology*, *9*(2), e1002927.

Lara, A. H., & Wallis, J. D. (2012). Capacity and precision in an animal model of short-term memory. *Journal of Vision*, *12*(3):13, 1–12, doi:10.1167/12.3.13. [PubMed] [Article]

Lee, D., & Chun, M. (2001). What are the units of visual short-term memory, objects or spatial locations? *Perception & Psychophysics*, *63*(2), 253–257.

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

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

Luria, R., & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia*, *49*(6), 1632–1639.

Ma, W. J., & Huang, W. (2009). No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint. *Journal of Vision*, *9*(11):3, 1–30, doi:10.1167/9.11.3. [PubMed] [Article]

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

Marshall, L., & Bays, P. M. (2013). Obligatory encoding of task-irrelevant features depletes working memory resources. *Journal of Vision*, *13*(2):21, 1–13, doi:10.1167/13.2.21. [PubMed] [Article]

Nilsson, T. H., & Nelson, T. M. (1981). Delayed monochromatic hue matches indicate characteristics of visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(1), 141–150.

Nolte, L. W., & Jaarsma, D. (1967). More on the detection of one of *M* orthogonal signals. *Journal of the Acoustical Society of America*, *41*(2), 497–505.

Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory and Cognition*, *41*, 1212–1227.

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*(1), 21–59.

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

Pearson, B., Raskevicius, J., Bays, P. M., Pertzov, Y., & Husain, M. (2014). Working memory retrieval as a decision process. *Journal of Vision*, *14*(2):2, 1–15, doi:10.1167/14.2.2. [PubMed] [Article]

Pelli, D. (1981). *The effects of visual noise*. Cambridge, UK: Cambridge University.

Shen, M., Tang, N., Wu, F., Shui, R., & Gao, Z. (2013). Robust object-based encoding in visual working memory. *Journal of Vision*, *13*(2):1, 1–11, doi:10.1167/13.2.1. [PubMed] [Article]

Shin, H., & Ma, W. J. (2016). Crowdsourced single-trial probes of visual working memory for irrelevant features. *Journal of Vision*, *16*(5):10, 1–8, doi:10.1167/16.5.10. [PubMed] [Article]

Sims, C. R. (2015). The cost of misremembering: Inferring the loss function in visual working memory. *Journal of Vision*, *15*(3):2, 1–27, doi:10.1167/15.3.2. [PubMed] [Article]

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal-observer analysis of visual working memory. *Psychological Review*, *119*(4), 807–830.

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

Van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding

precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences, USA, 109*(22), 8780–8785.

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron, 31*(4), 523–535.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 92–114.

Wheeler, M., & Treisman, A. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology General, 131*(1), 48–64.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*(12):11, 1120–1135, doi:10.1167/4.12.11. [PubMed] [Article]

Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review, 15*(1), 223–229.

Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 458–468.

Xu, Y. (2010). The neural fate of task-irrelevant features in object-based processing. *Journal of Neuroscience, 30*(42), 14020–14028.

Yin, J. J., Zhou, J., Xu, H., Liang, J., Gao, Z., & Shen, M. (2012). Does high memory load kick task-irrelevant information out of visual working memory? *Psychonomic Bulletin and Review, 19*(2), 218–224.

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

Appendix A: Derivation of decision rules

Here we outline the derivation of the decision rules (Step 2 in the Bayesian model) for Conditions A and B. The rules for Conditions C and D are specified in the main text.

Condition A: One feature dimension

In the model, observers compute the posterior probability distribution $p(L|x,y)$ of change location L given the measurement vectors x and y, and choose the location for which this posterior is highest. Denoting by a binary variable C_L whether the change occurred at the Lth location, we compute the posterior as

$$\begin{aligned} p(L|\mathbf{x},\mathbf{y}) &\propto p(\mathbf{x},\mathbf{y}|L)p(L) \\ &\propto p(\mathbf{x},\mathbf{y}|L) \\ &= p(x_L,y_L|C_L=1)\prod_{i\neq L}p(x_i,y_i|C_i=0) \\ &\propto \frac{p(x_L,y_L|C_L=1)}{p(x_L,y_L|C_L=0)}. \end{aligned}$$

In other words, the posterior probability that the change occurred at the Lth location is proportional to the likelihood ratio of change occurrence at that location considered independently from all other locations. This likelihood ratio is our decision variable and denote it by d_L:

$$d_L \equiv \frac{p(x_L,y_L|C_L=1)}{p(x_L,y_L|C_L=0)} \quad (7)$$

We evaluate numerator and denominator by marginalizing over $\boldsymbol{\theta}$, $\boldsymbol{\varphi}$, and in the case of $C_L=1$, also Δ:

$$p(x_L,y_L|C_L=1) = \int\int p(x_L|\theta_L)p(y_L|\varphi_L = \theta_L+\Delta)d\theta_L d\Delta$$

$$p(x_L,y_L|C_L=0) = \int p(x_L|\theta_L)p(y_L|\varphi_L=\theta_L)d\theta_L$$

Using Equation 1, the end result is (Van den Berg et al., 2012)

$$d_L = \frac{I_0(\kappa_{x,L})I_0(\kappa_{y,L})}{I_0\left(\sqrt{\kappa_{x,L}^2 + \kappa_{y,L}^2 + 2\kappa_{x,L}\kappa_{y,L}\cos(x_L-y_L)}\right)}$$

$$(8)$$

The maximum a posteriori observer will pick the location L for which d_L is highest.

Condition B: Two relevant feature dimensions

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
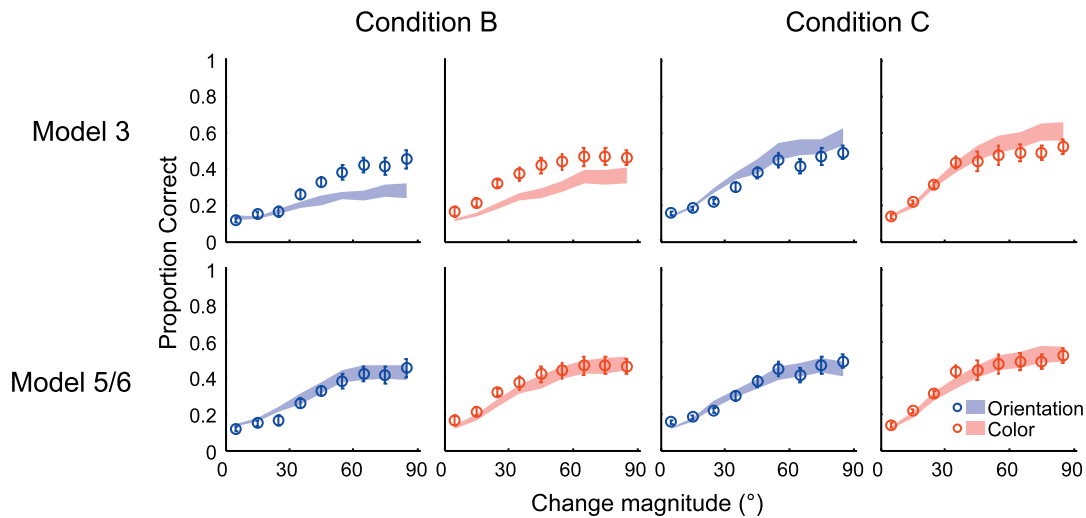
Figure A1. Replication of Experiment 2 with set size 8. Proportion correct as a function of change magnitude (five subjects), with model fits (shaded areas: $M \pm SEM$). Performance was indistinguishable between Conditions B and C, in accordance with Model 5/6 but not with Model 3.

changed and orientation has not. These hypotheses have the same a priori probability (0.5). Thus, Equation 7 for the decision variable changes to:

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

This can be simplified to

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

where

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

and analogous for color. Again, the maximum a posteriori observer will pick the location L for which d_L is highest.

Appendix B: Replication of Experiment 2

We replicated Experiment 2 at a higher set size ($N = 8$). Each display contained eight colored ellipses, equally spaced around an imaginary circle. Performance was similar between Conditions B and C (Figure A1). The AIC from Model 5/6 was higher than Model 3 (81 ± 19). This is consistent with Experiment 2.