Variability in encoding precision accounts for visual short-term memory limitations

Ronald van den Berg[a,1], Hongsup Shin[a,1], Wen-Chuang Chou[a,2], Ryan George[a,b], and Wei Ji Ma[a,3]

[a]Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030; and [b]Department of Computational and Applied Mathematics, Rice University, Houston TX 77005

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved April 11, 2012 (received for review October 24, 2011)

It is commonly believed that visual short-term memory (VSTM) consists of a fixed number of "slots" in which items can be stored. An alternative theory in which memory resource is a continuous quantity distributed over all items seems to be refuted by the appearance of guessing in human responses. Here, we introduce a model in which resource is not only continuous but also variable across items and trials, causing random fluctuations in encoding precision. We tested this model against previous models using two VSTM paradigms and two feature dimensions. Our model accurately accounts for all aspects of the data, including apparent guessing, and outperforms slot models in formal model comparison. At the neural level, variability in precision might correspond to variability in neural population gain and doubly stochastic stimulus representation. Our results suggest that VSTM resource is continuous and variable rather than discrete and fixed and might explain why subjective experience of VSTM is not all or none.

working memory | Bayesian inference | attention | estimation | change localization

Thomas Chamberlin famously warned scientists against entertaining only a single hypothesis, for such a modus operandi might lead to undue attachment and "a pressing of the facts to make them fit the theory" (ref. 1, p. 840). For half a century, the study of short-term memory limitations has been dominated by a single hypothesis, namely that a fixed number of items can be held in memory and any excess items are discarded (2–5). The alternative notion that short-term memory resource is a continuous quantity distributed over all items, with a lower amount per item translating into lower encoding precision, has enjoyed some success (6–8), but has been unable to account for the finding that humans often seem to make a random guess when asked to report the identity of one of a set of remembered items, especially when many items are present (9). Specifically, if resource were evenly distributed across items (6, 10), observers would never guess. Thus, at present, no viable continuous-resource model exists.

Here, we propose a more sophisticated continuous-resource model, the variable-precision (VP) model, in which the amount of resource an item receives, and thus its encoding precision, varies randomly across items and trials and on average decreases with set size. Resource might correspond to the gain of a neural population pattern of activity encoding a memorized feature. When gain is higher, a stimulus is encoded with higher precision (11, 12). Variability in gain across items and trials is consistent with observations of single-neuron firing rate variability (13–15) and attentional fluctuations (16, 17).

We tested the VP model against three alternative models (Fig. 1). According to the classic item-limit (IL) model (4), a fixed number of items is kept in memory, and memorized items are recalled perfectly. In the equal-precision (EP) model (6, 10), a continuous resource is evenly distributed across all items. The slots-plus-averaging (SA) model (9) acknowledges the presence of noise but combines it with the notion of discrete slots. Resource consists of a few discrete chunks, each of which affords limited precision to the encoding of an item. When there are fewer items than chunks, an item might get encoded using multiple chunks and thus with higher precision. To compare the four models, we used two visual short-term memory (VSTM) paradigms, namely delayed estimation (7) and change localization, each of which we applied to two feature dimensions, color and orientation (Fig. 2). We found that the VP model outperforms the previous models in each of the four experiments and accounts, at each set size, for the frequency that observers appear to be guessing. Thus, the VP model poses a serious challenge to models in which VSTM resource is assumed to be discrete and fixed.

Theory

VSTM Encoding and Variable Precision. An observer memorizes N simultaneously presented stimuli. The task-relevant feature is orientation or color, both of which are circular variables in our experiments. Each stimulus is encoded with precision J, which is formally defined as Fisher information (18). We assume that the observer's internal measurement of a stimulus is noisy and follows a Von Mises (circular normal) distribution,

$$p(x \mid s, J) = \mathrm{VM}\big(x; s, \kappa(J)\big) \equiv \frac{1}{2\pi I_0\big(\kappa(J)\big)} e^{\kappa(J)\cos(x-s)}, \qquad [1]$$

where I_0 is the modified Bessel function of the first kind of order 0 and the concentration parameter κ is uniquely determined by J through $J = \kappa \frac{I_1(\kappa)}{I_0(\kappa)}$ (*SI Text*). For a variable with a Gaussian distribution, J would be equal to inverse variance. A higher J produces a narrower distribution $p(x \mid s, J)$ (Fig. 3*A*). In the VP model, J is variable across items and trials and we assume that it is drawn, independently across items and trials, from a gamma distribution with mean $\bar{J}$ and scale parameter τ (Fig. 3*A*). The measurement is then described by a doubly stochastic process, $(\bar{J}, \tau) \to J \to x$. We further assume that $\bar{J}$ depends on set size, N, in power-law fashion, $\bar{J} = \bar{J}_1 N^{-\alpha}$ (Fig. 3*B*). The free parameters $\bar{J}_1$, α, and τ are fitted to subject data.

Models for Delayed Estimation. In experiments 1 and 2, observers estimated the value of a remembered stimulus (Fig. 2 *A* and *B*). The stimulus estimate, denoted $\hat{s}$, is equal to the measurement, x. In the IL model, the measurement of a remembered stimulus is noiseless but only K items (the "capacity") are remembered (or all N when $N \le K$), producing a guessing rate of $1 - K/N$ for $N > K$. In the SA model, K chunks of resource are allocated and the estimate distribution has two components. When the tested item

Author contributions: R.v.d.B., H.S., and W.J.M. designed research; R.v.d.B., H.S., W.-C.C., R.G., and W.J.M. performed research; R.v.d.B., H.S., W.-C.C., and R.G. analyzed data; and R.v.d.B. and W.J.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]R.v.d.B. and H.S. contributed equally to this work.

[2]Present address: Max Planck Institute for Dynamics and Self-Organization, Georg August University Göttingen, 37077 Göttingen, Germany.

[3]To whom correspondence should be addressed. E-mail: wjma@bcm.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1117465109/-/DCSupplemental.

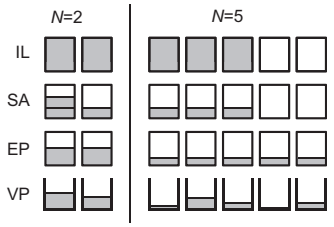

Fig. 1. Resource allocation in models of VSTM. Each box represents an item. Set size is 2 (*Left*) or 5 (*Right*). In this example, the number of "slots" or "chunks" is 3 in the IL and SA models.

has no chunks, the observer guesses and the estimate distribution is uniform; otherwise, it is a Von Mises distribution with κ determined by the number of chunks. In the EP model, the estimate distribution is Von Mises as in Eq. **1**, but with precision J equal across items and across trials with the same N and dependent on N as $J = J_1 N^{-\alpha}$. In the VP model, the estimate distribution is a mixture of many Von Mises distributions, each with a different value of κ: $p(\hat{s}\,|\,s) = \int \mathrm{VM}(\hat{s}; s, \kappa(J)) p(J\,|\,\bar{J}, \tau) dJ$ (Fig. S1*A*). In all models, we assume that the observer's response is equal to the estimate $\hat{s}$ plus zero-mean Von Mises response noise with concentration parameter κ_r. Model details can be found in *SI Text*.

Models for Change Localization. In experiments 3 and 4, observers sequentially viewed two displays, which were identical except that one stimulus changed between them. Observers reported where the change occurred (Fig. 2 *C* and *D*). The stimuli in the first display and the magnitude of the change were all drawn independently from a uniform distribution. In each model, stimuli are encoded in the same way as in delayed estimation, but the decision-making stage is different (Fig. 3*C*). We denote the measurements of the stimuli in the first and second displays by vectors $\mathbf{x}$ and $\mathbf{y}$, respectively, and the corresponding concentration parameters by a vector $\boldsymbol{\kappa}$. In the EP and VP models, the observer has access to all N pairs of measurements, but in the SA model only to K of them (or N when $N \leq K$). The statistical structure of the task-relevant variables is shown in Fig. S1*C*. In all models with noisy encoding, the observer's decision process is modeled as Bayesian inference. The Bayesian decision rule is to report the location L for which the posterior probability of change occurrence is largest, which is equivalent to the quantity $\frac{I_0(\kappa_L)^2}{I_0\left(\kappa_L\sqrt{2+2\cos(x_L - y_L)}\right)}$ being largest (*SI Text*).

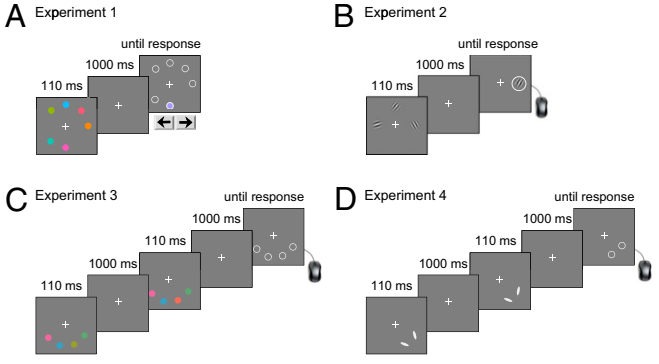

Fig. 2. Trial procedures. (*A*) Experiment 1: delayed estimation of color. Subjects scroll through all possible colors to report the remembered color in the marked location. (*B*) Experiment 2: delayed estimation of orientation. (*C*) Experiment 3: color change localization. (*D*) Experiment 4: orientation change localization.

Psychophysics and Model Comparison

Experiment 1: Delayed Estimation of Color. To compare the models, we first performed a delayed-estimation experiment (7). Observers briefly viewed and memorized the colors of N discs ($N = 1, \ldots, 8$) and reported the color of a randomly chosen target disk by scrolling through all possible colors (Fig. 2*A*). Following other authors (9), we fitted to the observer's estimation errors a mixture of a Von Mises distribution and a uniform distribution (see Fig. S2 for an example). We refer to the mixture proportion of the Von Mises component as w and to its circular SD as CSD. Note that this fitting procedure does not constitute a model, but is simply a way of summarizing the data into two descriptive statistics. It would be premature to interpret w as the probability that an item was encoded and $1 - w$ as the guessing rate, as suggested in ref. 9, because such an interpretation is meaningful only if the true error distribution is a uniform+Von Mises mixture, which we argue here is not the case. We verified that observers did not report colors of nontarget discs (Fig. S3; a different response modality, namely clicking on a color wheel, did produce nontarget reports). For each model, we generated synthetic datasets of the same size as the subject datasets, using the maximum-likelihood estimates of the parameters obtained from the subject data (Table S1), and then fitted the uniform+Von Mises mixture to these synthetic data. The resulting model predictions, averaged over subjects, are shown in Fig. 4*A* (for individual-subject fits, see Fig. S4). Consistent with previous results (9), we find a significant main effect of set size on both w [one-way repeated-measures ANOVA; $F(7, 84) = 42.1$, $P < 0.001$] and CSD [$F(7, 84) = 4.60$, $P < 0.001$]. This result rules out both the EP model, which predicts w close to 1 at each set size (the slight deviation is an artifact of the limited number of trials), and the IL model, which predicts that CSD is constant. The SA and VP models explain the data better, with the VP model having the lowest root mean-square (RMS) error (Fig. 4*A*). In the SA model, capacity K equals 4.00 ± 0.34 (mean $\pm$ SEM), in line with earlier work (9). In the VP model, the power α equals 1.33 ± 0.14 (Fig. S5*A*).

There is a clear intuition for why the VP model, but not the EP model, accounts for the decrease of w with set size. Because of trial-to-trial variability in precision, the target item sometimes, by chance, receives so little resource that the estimate on that trial is grouped into the uniform distribution, even though it was not a "real" guess. When set size is larger, mean precision is lower, resulting in more probability mass near zero precision (Fig. 3*B*) and a higher apparent guessing rate. Thus, it is not necessary to assume discrete resources to explain the decrease of w with set size.

To further determine which model best describes the data, we performed Bayesian model comparison (19), a principled method that automatically corrects for the number of free parameters (*SI Text*). We found that the log likelihood of the VP model exceeds those of the IL, SA, and EP models by respectively 15.6 ± 3.1, 12.0 ± 3.1, and 40.3 ± 6.3 points (Fig. 5*A*). A log-likelihood difference (or log Bayes factor) of 12.0 means that the data are $e^{12.0}$ times more probable under one model than under another. At the level of individual subjects (Fig. S6*A*), we find that the VP model is most likely for 12 of 13 subjects, whereas SA is slightly better for one. Consistent results were obtained using the Bayesian information criterion (20) (Fig. S6*B*).

Residual in Delayed Estimation. The VP model makes an intuitive prediction distinct from the other models. So far, we have fitted the data with a uniform+Von Mises mixture to obtain two descriptive statistics, w and CSD. The VP model postulates variability in precision, causing its predicted error distribution to be a mixture of a large number of Von Mises distributions, each with a different J. Such a mixture cannot be fitted perfectly with a uniform+Von Mises mixture and will therefore leave a residual.

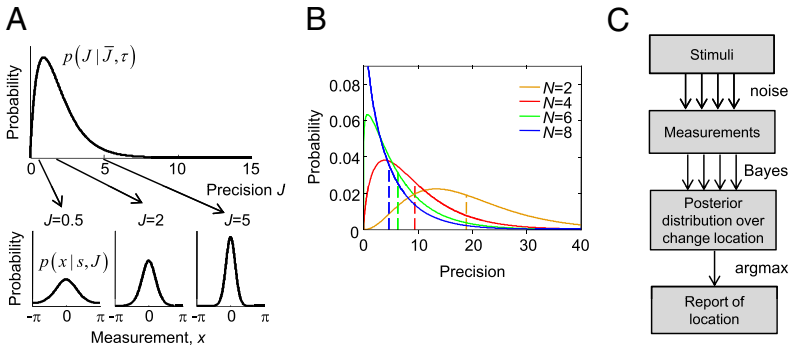

Fig. 3. Theory. (*A*) (*Upper*) In the VP model, precision, *J*, is variable and assumed to follow a gamma distribution (here with $\tau = 1$). (*Lower*) Von Mises noise distributions corresponding to three values of precision and $s = 0$. (*B*) Example probability distributions over precision at different 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. (*C*) Decision process in the Bayesian model of change localization.

Using the synthetic data described above, we find that the residual predicted by the VP model, but not by other models, has a central peak and negative side lobes (Fig. 5*B*). The subject data show a residual of exactly this shape (Fig. 5*C* and Fig. S2). This result constitutes additional evidence for variability in precision.

Experiment 2: Delayed Estimation of Orientation. To investigate the generality of these results, we replicated the experiment using orientation (Fig. 2*B*). The data show a significant main effect of set size on both *w* [one-way repeated-measures ANOVA, $F(7, 35) = 32.4$, $P < 0.001$] and CSD [$F(7, 35) = 3.28$, $P < 0.01$] (Fig. 4*B* and Fig. S7), again ruling out the IL and EP models. The SA and VP models explain the data better, with the VP model having the lowest RMS error (Fig. 4*B*). In the SA model, capacity $K = 3.33 \pm 0.56$. In the VP model, the power $\alpha = 1.41 \pm 0.15$ (Fig. S5*A*). Bayesian model comparison shows that the VP model outperforms the IL, SA, and EP models by 103 ± 15, 52 ± 11, and 142 ± 30 log-likelihood points, respectively (Fig. 5*D*). The VP model is most likely for all six subjects (Fig. S6*C*). Results were confirmed using the Bayesian information criterion (Fig. S6*D*). The residual after subtracting the uniform+Von Mises mixture has the shape predicted by the VP model (Fig. 5 *E* and *F*).

Experiments 3 and 4: Change Localization. To examine whether the VP model can account for human behavior in other VSTM tasks,

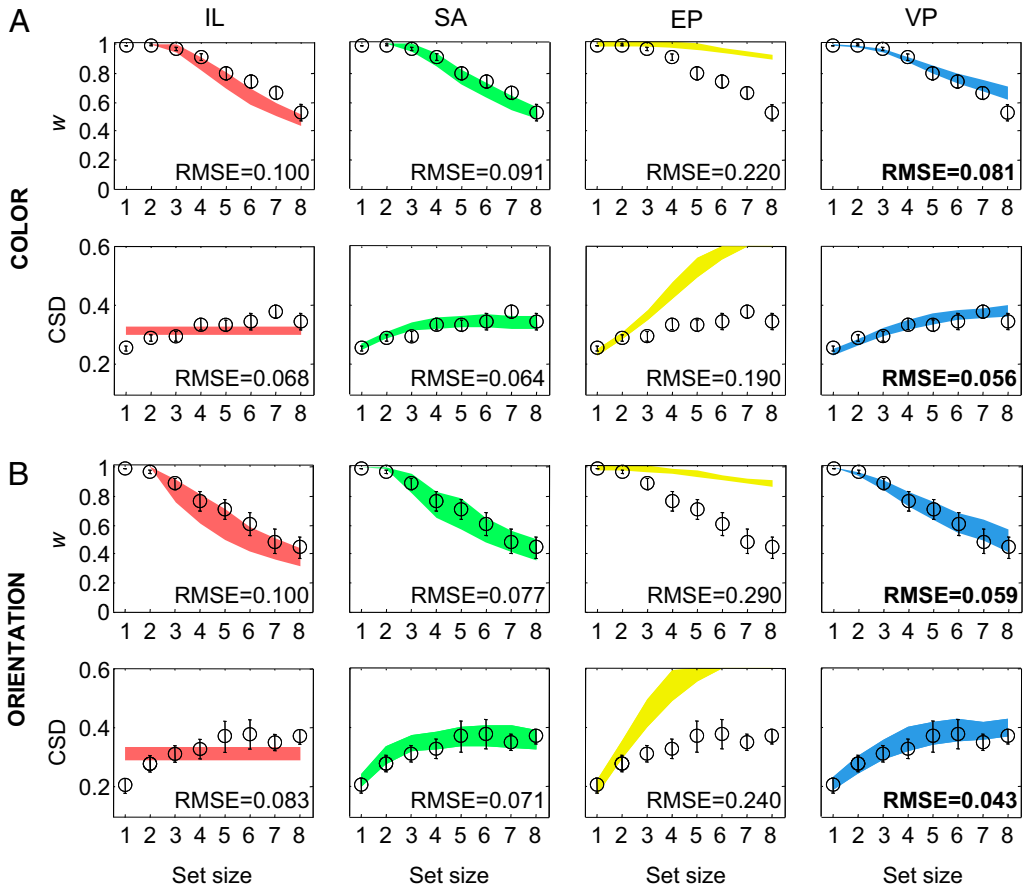

Fig. 4. (*A* and *B*) Parameters *w* and CSD obtained from fitting a mixture of a uniform and a Von Mises distribution to the estimation errors in experiment 1 (*A*) and experiment 2 (*B*). Here and elsewhere, circles and error bars represent data (mean and SEM) and shaded areas model predictions (SEM). Root mean-square error (RMSE) was computed across all set sizes and all subjects. The lowest RMSE in each comparison is indicated in boldface type.

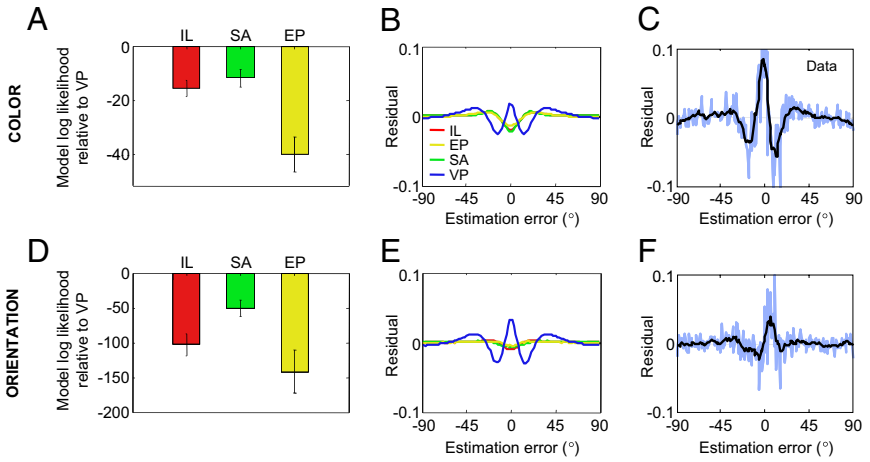

Fig. 5. More delayed-estimation results. (*A*) Model log likelihoods relative to the VP model in experiment 1 (colors). (*B*) Model predictions for the residual remaining after fitting a mixture of a uniform and a Von Mises distribution to the predicted error distribution, averaged over set sizes and subjects. (*C*) Blue: Residual after fitting a mixture of a uniform and a Von Mises distribution to the empirical error distribution. Black: Running average over a 0.28-rad window. (*D*–*F*) Same as *A*–*C*, but for experiment 2 (orientation).

we conducted two experiments in which subjects localized a change in the color or orientation of a stimulus (Fig. 2 *C* and *D*). 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 for orientation [$F(3, 30) = 356.5$, $P < 0.001$] (Figs. S8*A* and S9*A*). 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 for orientation [$F(8, 80) = 238.5$, $P < 0.001$] (Fig. 6). Judged by RMS error, the VP model provides the best fits to the psychometric curves (Fig. 6). Individual-subject fits are shown in Figs. S8 and S9. In the SA model, capacity $K = 2.86 \pm 0.14$ for color and 4.09 ± 0.39 for orientation. In the VP model, the power $\alpha = 0.974 \pm 0.090$ for color and 0.993 ± 0.075 for orientation (Fig. S5*B*). In Bayesian model comparison, the VP model outperforms the IL, SA, and EP models 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. 7 *A* and *C*). In both experiments, the VP model outperforms all other models for every individual subject (Fig. S10).

Apparent Guessing in Change Localization. To further distinguish the models, we computed an apparent guessing rate analogous to $1 - w$ in delayed estimation. We 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 the model-generated synthetic data. The EP model predicts an apparent guessing rate of zero. We 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. 7 *B* and *D*). 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.

Discussion

Do Slots Exist? Our results 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, we have shown that when

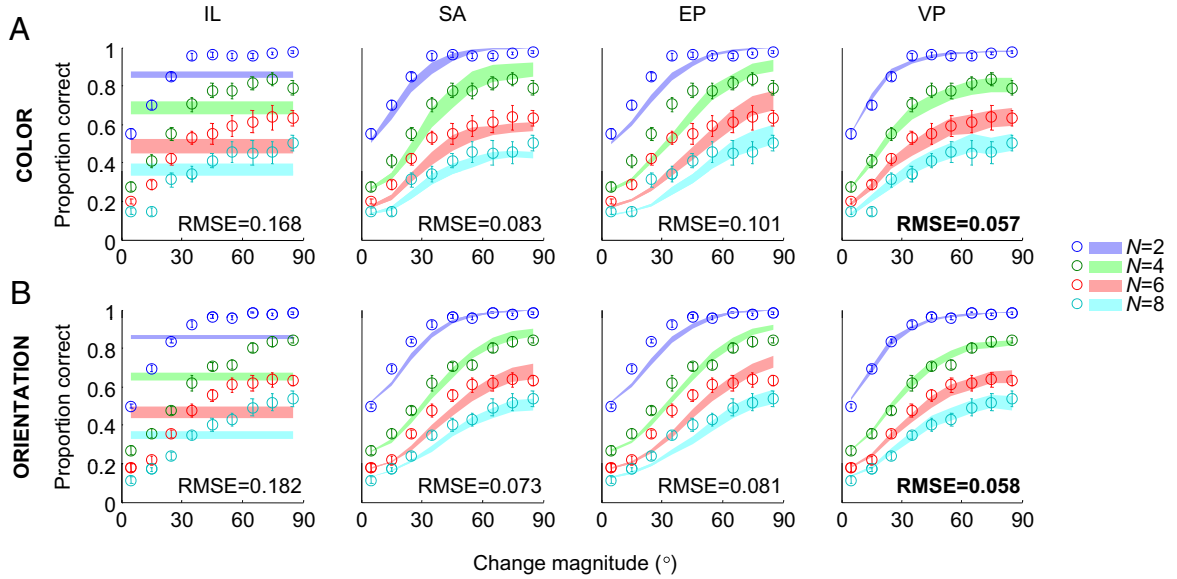

Fig. 6. (*A* and *B*) Proportion correct as a function of change magnitude at each set size in experiment 3 (*A*) and experiment 4 (*B*).

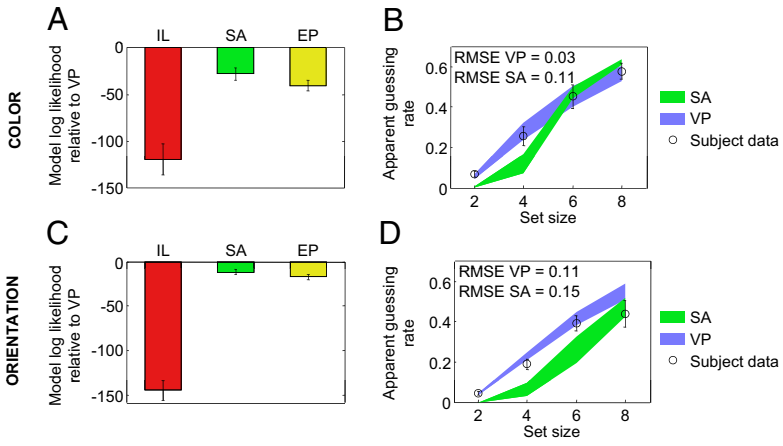

Fig. 7. More change localization results. (*A*) Model log likelihoods relative to the VP model in experiment 3 (colors). (*B*) Apparent guessing rate as a function of set size in experiment 3. (*C* and *D*) Same as *A* and *B*, but for experiment 4 (orientation).

such variability is not modeled, as in the EP model, human responses in delayed estimation and change localization cannot be accounted for. By contrast, the 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 poses a serious challenge to the notion of slots in VSTM and might reconcile an apparent capacity of about four items with the subjective sense that we possess some memory of an entire scene: Items are never discarded completely, but their encoding quality could by chance be very low.

Most neuroimaging and EEG studies of VSTM limitations consider only the slots framework (5, 21–24) (but see refs. 25 and 26). Without testing alternative models of VSTM, these studies cannot provide evidence for the existence of slots. The VP model offers a viable alternative, and we expect that quantities in the VP model will also correlate with neural variables.

We do not expect the VP model to end the debate about the nature of VSTM limitations. Variants of both the VP model and previous models can be conceived and should be tested. Possible hybrids between the SA and VP models include SA with trial-to-trial variability in capacity K (27, 28) and VP augmented with an item limit (continuous resource in discrete slots). We expect, however, that any alternative model will have to explicitly model variability in resource across items and trials to account for the data.

Is Resource Discrete? The 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, prepackaged boxes" (9) of resource, because those boxes allow for only a small, discrete number of noise levels. The VP model does not have this problem, because precision is a continuous quantity and is modulated by contrast in a continuous manner.

Neural Basis of VSTM Resource. Previous models have not specified a neural correlate of VSTM resource. Here, we 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 (29). Moreover, the energy cost associated with high gain (30) 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 (31–33), and attentional limitations are closely related to working memory ones (8, 34).

Neural Basis of Variability in Precision. Although our results point to variability in encoding precision as key in describing VSTM limitations, the VP model does not specify the origin of this variability. Variations in attention and alertness are likely contributors, but stimulus-related precision differences [such as cardinal orientations being encoded with higher precision (35)] might also play a role. There is evidence that microsaccades are predictive of variability in precision during change detection (36). Variability in precision provides a behavioral counterpart to recent physiological findings of trial-to-trial and item-to-item fluctuations in attentional gain (16, 17). A consequence of gain variability is that the neural representation $\mathbf{r}$ of a stimulus follows a doubly stochastic process $(\bar{g}, \tau) \rightarrow g \rightarrow \mathbf{r}$: The spike count distribution is determined by gain g, which itself is stochastic. Supporting this notion, doubly stochastic processes can well describe spike counts in lateral intraparietal cortex (LIP) (13), visual cortex (15), and other areas (14). Thus, the VP model is broadly consistent with emerging physiological findings.

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 LIP, an area associated with spatial attention, are lower to the onset of four than to that of two choice targets (37). In the superior colliculus, an area associated with covert attention, firing rates also decrease with the number of choice targets (38). Similar measurements in areas encoding short-term memories of visual stimuli remain to be made.

In both change localization experiments, we 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, in both delayed-estimation experiments, we found a steeper decline. This result 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 (39).

Neural Decoding. Nonhuman primate studies have begun to investigate set size effects in VSTM (36, 40–42). Advances in

simultaneous recordings from large populations of single neurons, as well as in the decoding of voxel patterns in functional MRI, might soon allow for model comparison more powerful than psychophysics allows. For instance, in delayed estimation, one could conceivably obtain estimates $\mathbf{x} = (x_1, \ldots, x_N)$ of the stimuli $\mathbf{s} = (s_1, \ldots, s_N)$ at all N locations simultaneously. The predictions for $p(\mathbf{x} \mid \mathbf{s})$ made by the SA and VP models can then be compared directly. Altogether, the VP model could help to consolidate the perspectives of cognitive psychology and systems neuroscience on VSTM limitations.

Methods

Detailed experimental methods can be found in *SI Text*. In experiment 1 (Fig. 2*A*), observers memorized the colors of N discs (N = 1, ... , 8) and reported the color of a randomly chosen target disk. Data of one subject were excluded, because her estimated value of w at set size 1 was extremely low (w = 0.72, compared with $w > 0.97$ for every other subject). A trial sequence consisted of the presentation of a fixation cross, the stimulus array, a delay period, and a response screen. Subjects responded by scrolling through all possible colors. Colors were drawn independently from a uniform distribution on a color wheel. Fourteen subjects each completed 864 trials in the scrolling condition. Experiment 2 (Fig. 2*B*) was identical except that stimuli were oriented Gabors. Set size was 2, 4, 6, or 8. Six subjects each completed 2,560 trials. In experiment 3 (Fig. 2*C*), observers were presented briefly with two displays containing N colored discs each (N = 2, 4, 6, or 8). The trial sequence consisted of the presentation of a fixation cross, the first stimulus array, a delay period, the second stimulus array, in which exactly one stimulus had changed color, and a response screen. Subjects clicked on the location of the stimulus that had changed. Colors in the first array and the magnitude of the change were drawn independently from a uniform distribution on a color wheel. Seven subjects each completed 1,920 trials. Experiment 4 (Fig. 2*D*) was identical except that stimuli were oriented ellipses. Eleven subjects each completed 1,920 trials.

Data Analysis. We used maximum-likelihood fitting and Bayesian model comparison. We verified numerical robustness (Fig. S11). All methods are discussed in *SI Text*.

ACKNOWLEDGMENTS. W.J.M. is supported by Award R01EY020958 from the National Eye Institute. R.v.d.B. was supported by the Netherlands Organisation for Scientific Research.

1. Chamberlin TC (1897) The method of multiple working hypotheses. *J Geol* 5:837–848.
2. Miller GA (1956) The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychol Rev* 63:81–97.
3. Cowan N (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav Brain Sci* 24:87–114, discussion 114–185.
4. Pashler H (1988) Familiarity and visual change detection. *Percept Psychophys* 44: 369–378.
5. Fukuda K, Awh E, Vogel EK (2010) Discrete capacity limits in visual working memory. *Curr Opin Neurobiol* 20:177–182.
6. Palmer J (1990) Attentional limits on the perception and memory of visual information. *J Exp Psychol Hum Percept Perform* 16:332–350.
7. Wilken P, Ma WJ (2004) A detection theory account of change detection. *J Vis* 4: 1120–1135.
8. Bays PM, Husain M (2008) Dynamic shifts of limited working memory resources in human vision. *Science* 321:851–854.
9. Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. *Nature* 453:233–235.
10. Shaw ML (1980) Identifying attentional and decision-making components in information processing. *Attention and Performance*, ed Nickerson RS (Erlbaum, Hillsdale, NJ), Vol VIII, pp 277–296.
11. Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. *Proc Natl Acad Sci USA* 90:10749–10753.
12. Knill DC, Pouget A (2004) The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci* 27:712–719.
13. Churchland AK, et al. (2011) Variance as a signature of neural computations during decision making. *Neuron* 69:818–831.
14. Churchland MM, et al. (2010) Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nat Neurosci* 13:369–378.
15. Goris RLT, Simoncelli EP, Movshon JA (2012) Using a doubly-stochastic model to analyze neuronal activity in the visual cortex. *Cosyne Abstracts* (Salt Lake City).
16. Cohen MR, Maunsell JHR (2010) A neuronal population measure of attention predicts behavioral performance on individual trials. *J Neurosci* 30:15241–15253.
17. Nienborg H, Cumming BG (2009) Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature* 459:89–92.
18. Cover TM, Thomas JA (1991) *Elements of Information Theory* (John Wiley & Sons, New York).
19. MacKay DJ (2003) *Information Theory, Inference, and Learning Algorithms* (Cambridge Univ Press, Cambridge, UK).
20. Schwartz GE (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464.
21. Anderson DE, Vogel EK, Awh E (2011) Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *J Neurosci* 31:1128–1138.
22. Todd JJ, Marois R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428:751–754.
23. Vogel EK, Machizawa MG (2004) Neural activity predicts individual differences in visual working memory capacity. *Nature* 428:748–751.
24. Sauseng P, et al. (2009) Brain oscillatory substrates of visual short-term memory capacity. *Curr Biol* 19:1846–1852.
25. Magen H, Emmanouil T-A, McMains SA, Kastner S, Treisman A (2009) Attentional demands predict short-term memory load response in posterior parietal cortex. *Neuropsychologia* 47:1790–1798.
26. Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440:91–95.
27. Dyrholm M, Kyllingsbaek S, Espeseth T, Bundesen C (2011) Generalizing parametric models by introducing trial-by-trial parameter variability: The case of TVA. *J Math Psych* 55:416–429.
28. Sims CR, Jacobs RA, Knill DC (2012) An ideal-observer analysis of visual working memory. *Psychol Rev*, in press.
29. Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9:1432–1438.
30. Lennie P (2003) The cost of cortical computation. *Curr Biol* 13:493–497.
31. McAdams CJ, Maunsell JH (1999) Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron* 23:765–773.
32. Treue S, Martínez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575–579.
33. Salinas E, Sejnowski TJ (2001) Gain modulation in the central nervous system: Where behavior, neurophysiology, and computation meet. *Neuroscientist* 7:430–440.
34. Awh E, Jonides J (2001) Overlapping mechanisms of attention and spatial working memory. *Trends Cogn Sci* 5:119–126.
35. Girshick AR, Landy MS, Simoncelli EP (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat Neurosci* 14:926–932.
36. Lara AH, Wallis JD (2012) Capacity and precision in an animal model of short-term memory. *J Vis* 12:1–12.
37. Churchland AK, Kiani R, Shadlen MN (2008) Decision-making with multiple alternatives. *Nat Neurosci* 11:693–702.
38. Basso MA, Wurtz RH (1998) Modulation of neuronal activity in superior colliculus by changes in target probability. *J Neurosci* 18:7519–7534.
39. Mazyar H, Van den Berg R, Ma WJ (2012) Does precision decrease with set size? *J Vis*, in press.
40. Heyselaar E, Johnston K, Pare M (2011) A change detection approach to study visual working memory of the macaque monkey. *J Vis* 11(3):11, 1–10.
41. Elmore LC, et al. (2011) Visual short-term memory compared in rhesus monkeys and humans. *Curr Biol* 21:975–979.
42. Buschman TJ, Siegel M, Roy JE, Miller EK (2011) Neural substrates of cognitive capacity limitations. *Proc Natl Acad Sci USA* 108:11252–11255.

Supporting Information

Supporting Information Corrected December 17, 2012

Van den Berg et al. 10.1073/pnas.1117465109

SI Text

Convention: For convenience, we have mapped both orientation and color space to $[0, 2\pi)$ in all equations.

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

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 (1, 2). Each observation is corrupted by independent, zero-mean Gaussian noise with the same SD, and the observer's eventual measurement, x, is the mean of these observations. Then the variance of the measurement decreases inverse proportionally to the number of observations, and precision increases proportionally.

In this paper, we instead identify resource with Fisher information, denoted J. Fisher information determines the best possible performance of any estimator, through the Cramér-Rao bound (3). Fisher information is defined in terms of the noise distribution, which is the distribution of the observations conditioned on the stimulus s,

$$J(s) = -\left\langle \frac{\partial^2}{\partial s^2} \log p(\text{observations}\,|\,s) \right\rangle, \qquad \textbf{[S1]}$$

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

If x follows a Gaussian distribution with mean s and SD σ, it is easily verified from the definition, Eq. **S1**, that Fisher information is equal to the inverse variance, $J = \frac{1}{\sigma^2}$, recovering 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 (4).

A slight complication arises from the fact that the stimulus spaces we use (orientation and color) are circular, so that the Gaussian distribution is no longer appropriate. Instead, we assume that the measurement follows a Von Mises distribution:

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

I_0 is the modified Bessel function of the first kind of order zero (5) 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. It is important that the Gaussian distribution is a special case of the Von Mises distribution, because the maximum-likelihood estimate has an asymptotically Gaussian distribution.

We calculate Fisher information from its definition, Eq. **S1**,

$$J = \langle \kappa \cos(x-s) \rangle = \frac{\kappa}{2\pi I_0(\kappa)} \int \cos(x-s) e^{\kappa \cos(x-s)} d\hat{s} = \kappa \frac{I_1(\kappa)}{I_0(\kappa)}, \qquad \textbf{[S2]}$$

where $I_1(\kappa)$ is the modified Bessel function of the first kind of order one (5). This equation relates Fisher information in a one-to-one fashion to the concentration parameter of the Von Mises distribution. We use it in all models except for the IL model. One can think of Fisher information as precision, by analogy to the Gaussian case. We write the inverse relationship of Eq. **S2** as

$$\kappa = \Phi(J). \qquad \textbf{[S3]}$$

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

Equal-precision model. In the equal-precision (EP) model, we assume

$$J = \frac{J_1}{N^\alpha}, \qquad \textbf{[S4]}$$

where J_1 is the Fisher information at set size 1. Using Eq. **S3**, the concentration parameter at set size N is

$$\kappa(N) = \Phi\left(\frac{J_1}{N^\alpha}\right). \qquad \textbf{[S5]}$$

Slots-plus-averaging model. The slots-plus-averaging (SA) model (6) 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 (6) do not mention the exact relationship between amount of resource and the concentration parameter of the Von Mises distribution, but we assume that they used the correct relationship, Eq. **S2**. Then, the concentration parameter as a function of S is

$$\kappa = \Phi(SJ_1), \qquad \textbf{[S6]}$$

where J_1 is now the Fisher information corresponding to having one chunk ($S = 1$). When $N > K$, an item receives 0 chunks or 1 chunk, with probabilities K/N and $1 - K/N$, respectively. This allocation is the same as in the IL model. When $N \le K$, 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. **S6**, these two values of S correspond to two values of the concentration parameter κ, which we denote by κ_{low} and κ_{high}, respectively:

$$\begin{aligned} \kappa_{\text{low}}(N) &= \Phi\left(\left\lfloor \frac{K}{N} \right\rfloor J_1\right), \\ \kappa_{\text{high}}(N) &= \Phi\left(\left(\left\lfloor \frac{K}{N} \right\rfloor + 1\right) J_1\right). \end{aligned} \qquad \textbf{[S7]}$$

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

Variable-precision model. In the variable-precision model, precision is variable across items and trials. We assume that each precision is drawn independently from a gamma distribution with mean precision $\bar{J}$ and scale parameter τ,

$$p(g \mid \bar{g}; \tau) = \text{Gamma}(J; \bar{J}, \tau). \quad \textbf{[S8]}$$

The variance of J is equal to $\bar{J}\tau$. The gamma distribution is a common distribution on the positive real line. We assume that mean precision depends on set size in the following way:

$$\bar{J} = \frac{\bar{J}_1}{N^{\alpha}}. \quad \textbf{[S9]}$$

Model Predictions for Delayed Estimation. ***Item-limit model.*** The item-limit model assumes that the memory of a stored item is perfect; thus $\hat{s} = s$. However, we allow for the possibility that response noise (e.g., motor noise) corrupts the subject's response. Therefore, we assume that the response, denoted r, follows a Von Mises distribution centered on the true stimulus with concentration parameter κ_r. For $N \leq K$, we then have $p(r \mid s) = \text{VM}(r; s, \kappa_r)$. If $N > K$, there is a probability of K/N that the probed item was memorized and a probability of $1 - K/N$ that it was not memorized, in which case the subject will make a random guess. Hence, the response distribution is a mixture of a Von Mises distribution and a uniform (guessing) distribution:

$$p(r \mid s) = \frac{K}{N}\text{VM}(r; s, \kappa_r) + \left(1 - \frac{K}{N}\right)\frac{1}{2\pi}. \quad \textbf{[S10]}$$

This model has two free parameters: K and κ_r.

Equal-precision model. In the presence of encoding noise, the best estimate of the stimulus is equal to the measurement, $\hat{s} = x$. The estimate distribution predicted by the EP model is then

$$p(\hat{s} \mid s) = \text{VM}(\hat{s}; s, \kappa(N))$$

with $\kappa(N) = \Phi\left(\frac{J_1}{N^{\alpha}}\right)$ (Eq. **S5**). Including response noise with concentration parameter κ_r, the response distribution is

$$p(r \mid s) = \int_0^{2\pi} \frac{1}{2\pi I_0(\kappa(N))} e^{\kappa(N)\cos(\hat{s}-s)} \frac{1}{2\pi I_0(\kappa_r)} e^{\kappa_r \cos(r-\hat{s})} d\hat{s}.$$

A lengthy but straightforward calculation gives

$$p(r \mid s; N) = \frac{I_0\left(\sqrt{\kappa(N)^2 + \kappa_r^2 + 2\kappa(N)\kappa_r \cos(r-s)}\right)}{2\pi I_0(\kappa_r) I_0(\kappa(N))}. \quad \textbf{[S11]}$$

The EP model for the delayed-estimation task has three free parameters: J_1, α, and κ_r.

Slots-plus-averaging model. The estimate distribution is a mixture of a Von Mises and a uniform distribution,

$$p(\hat{s} \mid s) = \frac{K}{N}\text{VM}(\hat{s}; s, \kappa_1) + \left(1 - \frac{K}{N}\right)\frac{1}{2\pi},$$

with $\kappa_1 = \Phi(J_1)$. With response noise, the response distribution becomes

$$p(r \mid s) = \frac{K}{N}\frac{I_0\left(\sqrt{\kappa_1^2 + \kappa_r^2 + 2\kappa_1\kappa_r \cos(r-s)}\right)}{2\pi I_0(\kappa_1) I_0(\kappa_r)} + \left(1 - \frac{K}{N}\right)\frac{1}{2\pi}. \quad \textbf{[S12]}$$

The estimate distribution for $N \leq K$ is a mixture of two Von Mises distributions:

$$p(\hat{s} \mid s) = \frac{K \bmod N}{N}\frac{1}{2\pi I_0(\kappa_{\text{high}}(N))} e^{\kappa_{\text{high}}(N)\cos(\hat{s}-s)} + \left(1 - \frac{K \bmod N}{N}\right)\frac{1}{2\pi I_0(\kappa_{\text{low}}(N))} e^{\kappa_{\text{low}}(N)\cos(\hat{s}-s)}.$$

With response noise, the response distribution for $N \leq K$ is

$$p(r \mid s) = \frac{K \bmod N}{N}\frac{I_0(\kappa_{c,\text{high}}(N))}{2\pi I_0(\kappa_{\text{high}}(N)) I_0(\kappa_r)} + \left(1 - \frac{K \bmod N}{N}\right)\frac{I_0(\kappa_{c,\text{low}}(N))}{2\pi I_0(\kappa_{\text{low}}(N)) I_0(\kappa_r)}, \quad \textbf{[S13]}$$

with

$$\kappa_{c,\text{high}}(N) = \sqrt{\kappa_{\text{high}}(N)^2 + \kappa_r^2 + 2\kappa_{\text{high}}(N)\kappa_r \cos(\hat{s}-s)},$$
$$\kappa_{c,\text{low}}(N) = \sqrt{\kappa_{\text{low}}(N)^2 + \kappa_r^2 + 2\kappa_{\text{low}}(N)\kappa_r \cos(\hat{s}-s)}. \quad \textbf{[S14]}$$

The SA model for the delayed-estimation task has three free parameters: K, J_1, and κ_r.

Variable-precision model. The estimate distribution corresponding to a fixed precision J is $p(\hat{s} \mid s; J) = \text{VM}(\hat{s}; s, \Phi(J))$. When precision is variable, the estimate distribution is a mixture of the estimate distributions associated with individual values of precision, with mixture proportions equal to the frequencies of those values, $p(J \mid \bar{J}; \tau) = \text{Gamma}(J; \bar{J}, \tau)$, with $\bar{J}$ given by Eq. **S9**. Therefore,

$$p(\hat{s} \mid s; \bar{J}, \tau) = \int p(\hat{s} \mid s; J) p(J \mid \bar{J}; \tau) dJ = \int \text{VM}(\hat{s}; s, \Phi(J))\,\text{Gamma}(J; \bar{J}, \tau) dJ, \quad \textbf{[S15]}$$

This distribution 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. We examined the effect of the number of samples on the model predictions and found that 500 is a sufficiently large number to give robust results (Fig. S11*A*). Response noise is added as above, by convolving $p(\hat{s} \mid s; \bar{J}, \tau)$ with a Von Mises distribution with concentration parameter κ_r. Thus, the variable-precision (VP) model for the delayed-estimation task has four free parameters: $\bar{J}_1$, α, τ, and κ_r.

Model Predictions for Change Localization. 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, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$, and the vector of stimuli in the first display, $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_N)$. Each L has a probability of $1/N$. The probability density of Δ is flat at $\frac{1}{2\pi}$, and the one over $\boldsymbol{\theta}$ is 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 zeros with a 1 at the Lth entry.

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 $N > K$. These probabilities are independent of $\boldsymbol{\theta}$, $\boldsymbol{\varphi}$, and Δ. We introduced ε because without it (i.e., $\varepsilon = 0$), the data would have probability zero under the model. The IL model for the change localization task has two free parameters: K and ε.

Bayesian decision rule. 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 ith 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 ith 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 we chose not to do so). The relations between the variables are shown

in the graphical model in Fig. S1*B*. To model how the observer decides on the basis of the measurements $\mathbf{x} = (x_1, \ldots, x_N)$ and $\mathbf{y} = (y_1, \ldots, y_N)$, we use a Bayesian-observer model. The Bayesian observer computes a probability distribution over the location of the change, $p(L \mid \mathbf{x}, \mathbf{y})$, 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 $\boldsymbol{\varphi}$,

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

where in going from the first to the second line we have used the structure of the generative model in Fig. S1*B*. Substituting distributions and evaluating the integral over $\boldsymbol{\varphi}$ gives

$$p(\mathbf{x},\mathbf{y},L) = \frac{1}{N}\left(\frac{1}{2\pi}\right)^{N+1} \int \prod_{i=1}^{N}\left(\int p(x_i\mid\theta_i)p(y_i\mid\varphi_i=\theta_i+\Delta\delta_{L,i})\right)d\Delta, \quad \textbf{[S16]}$$

where $\delta_{L,i} = 1$ when $L = i$ and 0 otherwise. Because we are interested only in the dependence on L, we can freely divide by the L-independent product $\prod_{i=1}^{N}(\int p(x_i\mid\theta_i)p(y_i\mid\varphi_i=\theta_i))$, leaving only integrals pertaining to the Lth location:

$$p(\mathbf{x},\mathbf{y},L) \propto \frac{\iint p(x_L\mid\theta_L)p(y_i\mid\varphi_L=\theta_L+\Delta)d\theta_L d\Delta}{\int p(x_L\mid\theta_L)p(y_L\mid\varphi_L=\theta_L)}. \quad \textbf{[S17]}$$

that is among the encoded ones. In analogy to Eq. **S16**, this probability is

$$\begin{aligned} &(L \text{ encoded})\, p(\mathbf{x},\mathbf{y},L) = \frac{1}{N}\left(\frac{1}{2\pi}\right)^{K+1} \\ &\times \int \prod_{i=1}^{K}\left(\int p(x_i\mid\theta_i)p(y_i\mid\varphi_i=\theta_i+\Delta\delta_{L,i})\right)d\Delta. \end{aligned} \quad \textbf{[S19]}$$

Now we 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 probability is equal to

$$\begin{aligned} (L \text{ not encoded})\, p(\mathbf{x},\mathbf{y},L) &= \iint p(\mathbf{x},\mathbf{y},\boldsymbol{\theta},\boldsymbol{\varphi},L)d\boldsymbol{\theta} d\boldsymbol{\varphi} \\ &= \iint p(L)p(\boldsymbol{\theta})p(\boldsymbol{\varphi}\mid L,\boldsymbol{\theta})p(\mathbf{x}\mid\boldsymbol{\theta})p(\mathbf{y}\mid\boldsymbol{\varphi})d\boldsymbol{\theta} d\boldsymbol{\varphi} \\ &= \frac{1}{N}\left(\frac{1}{2\pi}\right)^{K}\prod_{i=1}^{K}\left(\int p(x_i\mid\theta_i)p(y_i\mid\varphi_i=\theta_i)\right). \end{aligned} \quad \textbf{[S20]}$$

As one would expect, this probability does not depend on L. Because we are interested only in the location L for which $p(\mathbf{x}, \mathbf{y}, L)$ is largest (i.e., the argmax), we divide both Eqs. **S19** and **S20** by Eq. **S20**. Then, in analogy to Eq. **S17**, we have to take the argmax of

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

This probability 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

$$\hat{L} = \underset{L}{\operatorname{argmax}} \frac{I_0(\kappa_L)^2}{I_0\left(\kappa_L\sqrt{2+2\cos(x_L-y_L)}\right)}. \quad \textbf{[S18]}$$

The distribution of the MAP estimate for given L, Δ, N, and $\boldsymbol{\kappa}$, denoted $p(\hat{L}\mid L,\Delta,\boldsymbol{\kappa};N)$, depends on the model but is computed through Monte Carlo simulation for all models (using 10,000 samples of $\mathbf{x}$ and $\mathbf{y}$). Note that the estimate distribution is characterized by a single number, namely probability correct.

Equal-precision model. In the equal-precision model, we take $\kappa_i = \Phi\left(\frac{J_1}{N^{\alpha}}\right)$ for all i. We assume equality between first and second displays, because the concentration parameters in both displays can essentially not be fitted independently (compare with the sum of two normally distributed random variables: the variances sum and cannot be estimated individually). The EP model for the change localization task has two free parameters: J_1 and α.

Slots-plus-averaging model. In the slots-plus-averaging model, κ_i is given by Eq. **S7**. When $N \le K$, Eq. **S18** applies. When $N > K$, inference is based on only K of N measurements in each display, $\mathbf{x} = (x_1, \ldots, x_K)$ and $\mathbf{y} = (y_1, \ldots, y_K)$, yet the change could have occurred at any location. We first evaluate the joint probability of $\mathbf{x}$, $\mathbf{y}$ and that the change occurred at a location L

Evaluating the integral, the estimate of location is

$$\hat{L} = \underset{L}{\operatorname{argmax}} \frac{I_0(\kappa_L)^2}{I_0\left(\kappa_L\sqrt{2+2\cos(x_L-y_L)}\right)} \quad \textbf{[S21]}$$

when the value of this maximum exceeds 1, and we randomly guess from among the nonencoded items when it does not. The SA model for the change localization task has two free parameters: J_1 and K.

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. **S9**) and scale parameter τ. Then, the estimate distribution is

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

This distribution is obtained through Monte Carlo simulation of $\mathbf{J}$, using 10,000 samples. The VP model for the change localization task has three free parameters: $\bar{J}_1$, α, and τ.

Experimental Details. ***Experiment 1: Delayed estimation with color stimuli.*** Observers briefly viewed and memorized a set of colors and reported the color of a randomly chosen target disk (Fig. 2*A*).

Stimuli. Stimuli were displayed on a 21-inch cathode ray tube monitor at a viewing distance of ~60 cm. The stimulus array consisted of N colored discs ($N = 1, \ldots, 8$) with a diameter of 2° of visual angle, with their centers lying on an imaginary circle of radius 4.5°. 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 of the discs 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 ($L^* = 70$) and was centered at the point ($a^* = 10$, $b^* = 10$). The stimuli were presented on a midlevel gray background (128 on an 8-bit grayscale) of luminosity 8.1 cd/m^2.

Procedure. A trial sequence consisted of the presentation of a fixation cross, the stimulus array, and a delay period during which only the fixation cross was visible (Fig. 2*A*). Set size was chosen pseudorandomly and the colors of the items were drawn independently from a uniform distribution. The response screen consisted of white circles marking the circumferences of the discs in the stimulus array, with a thicker circle marking one randomly chosen disk. The subject's task was to report the color of the disk that had been present in the stimulus array at the marked location, by using the left and right arrow keys to scroll through all possible colors. After the first key press, a random color appeared within the thicker circle. After subsequent key presses, this color changed by moving either clockwise or counterclockwise through the color wheel. The association between left/right key presses and the direction in which the color wheel was traversed was randomized on each trial. To submit a response, the subject pressed the space bar.

The experiment consisted of three sessions on different days. Each session consisted of two blocks in which subjects responded using a color wheel condition (*SI Text*) and two blocks in which they responded by scrolling. Color wheel and scrolling blocks were interleaved in ABBA order, with A and B randomized for each subject. After every 24 trials, feedback was given in the form of a total score. The score per trial was 3 when the estimate was within five color values of the true value and was $\lfloor 3 - E/15 \rfloor$ (floor function), with E the error, otherwise. The first two blocks were each preceded by 8 practice trials. If the total score across these trials was less than 3, subjects were asked to repeat the practice. The actual block consisted of 144 trials. In total, each subject completed $3 \cdot 2 \cdot 2 \cdot 144 = 1{,}728$ testing trials.

Fourteen subjects participated in this experiment (age range, 18–50 y; 12 naive). Data of one subject were excluded, because her estimated value of w at set size 1 was extremely low ($w = 0.72$, compared with $w > 0.97$ for every other subject).

Experiment 2: Delayed estimation with orientation stimuli. Experiment 2 differed from experiment 1 in the following ways. Set size was 2, 4, 6, or 8. All stimuli were displayed on a 19-inch liquid crystal display monitor at a viewing distance of ∼50 cm. The stimulus array was composed of Gabors with a Gaussian envelope of 0.5° and a wavelength randomly chosen from a uniform distribution on [0.3, 0.8] cycles per degree (Fig. 2*B*). The Gabor centers lay on an imaginary circle of radius 8.2°. Presentation time was 110 ms. A circle appeared around the location of the item whose orientation had to be reported. When subjects moved the mouse, a Gabor appeared inside that circle. They had to rotate it using the mouse to match the orientation of the Gabor that had been in that location. They pressed the space bar to submit their response. Feedback consisted of an integer score between –3 and +3 on every trial. When E is the error, the score was computed as $3 - E/15$, rounded to the nearest integer. Six subjects participated (four naive). Each subject completed four sessions of 640 trials each, for a total of 2,560 trials.

Experiment 3: 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. 2*C*).

Stimuli. All stimuli were displayed on a 19-inch LCD monitor at a viewing distance of ∼60 cm. The first stimulus array was composed of N colored discs ($n = 2$, 4, 6, or 8) with a diameter of 0.62° of visual angle with their centers lying on an imaginary circle of radius 7° (Fig. 2*C*). 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 midlevel gray background (128 on an 8-bit grayscale) of luminosity 33.1 cd/m^2.

Procedure. The trial sequence consisted of the presentation of a fixation cross (1,000 ms), the stimulus array (110 ms), a delay period during which only the fixation cross was visible (1,000 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, and 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 four sessions on different days. Each session consisted of four blocks with 120 trials each. Hence, each subject completed $4 \cdot 4 \cdot 120 = 1{,}920$ trials in total. Seven subjects participated in this experiment (age range, 21–32 y; five naive).

Experiment 4: Change localization with orientation stimuli. Experiment 4 differed from experiment 3 in the following ways. Stimuli were white, oriented ellipses with minor and major axes of 0.41° and 0.94° of visual angle, respectively, and a luminance of 95.7 cd/m^2 (Fig. 2*D*). Eleven subjects participated (age range, 23–29 y; 9 naive).

Details of Data Analysis in Experiments 1 and 2. In experiments 1 and 2, to remove bias, we circularly subtracted, for each subject separately, the circular mean across all trials from the subject's reports before any analyses.

Computing the summary statistics w and CSD. In delayed estimation, the raw data consist of the distributions of the estimation error, Δs, at each set size (Fig. S2). The summary statistics w and CSD (Fig. 4) were obtained by fitting a mixture of a Von Mises distribution and a uniform distribution:

$$p_{\text{fit}}(\Delta s; w, \kappa_{\text{fit}}) = \frac{w}{2\pi I_0(\kappa_{\text{fit}})} e^{\kappa_{\text{fit}} \cos \Delta s} + \frac{1-w}{2\pi}. \qquad \textbf{[S22]}$$

The circular SD is defined as $\text{CSD} = \sqrt{1 - \frac{I_1(\kappa_{\text{fit}})}{I_0(\kappa_{\text{fit}})}}$ (7). We fitted this mixture separately for each subject and each set size, both to the data and to the error distributions predicted by each of the models. Fitting was done through maximum-likelihood estimation, which means choosing the values of the parameters of Eq. **S22**, w and κ_{fit}, that maximize the probability of the data given the parameters. This is equivalent to maximizing the log-likelihood function

$$\log L(w, \kappa_{\text{fit}}) = \log p(\text{data} \mid w, \kappa_{\text{fit}}) = \log \prod_{i=1}^{N_{\text{trials}}} p_{\text{fit}}(\Delta s_i \mid w, \kappa_{\text{fit}})$$

$$= \sum_{i=1}^{N_{\text{trials}}} \log p_{\text{fit}}(\Delta s_i \mid w, \kappa_{\text{fit}}),$$

where N_{trials} is the number of trials. We use fminsearch in Matlab to perform the maximization.

Nontarget reports in experiment 1. It has been argued that in the color wheel condition, guessing is confounded with nontarget reports, in the sense that the fitted uniform component includes a substantial amount of reports of nontarget colors (8). To test for this, we

fitted two modified mixtures to the data. The first is the one that assigns a probability to reporting the color of a nontarget disk (8),

$$p_{\text{fit}}(\hat{s}\,|\,s;w,\kappa_{\text{fit}}) = \frac{w_{\text{guess}}}{2\pi} + w\text{VM}(\hat{s},s,\kappa_{\text{fit}}) + (1-w-w_{\text{guess}})\,\frac{1}{N-1}\sum_{j=1}^{N-1}\text{VM}(\hat{s},s_j,\kappa_{\text{fit}}),$$

where $\hat{s}$ is the reported value, s is the target value, s_j is the jth nontarget value, w_{guess} is the guessing rate, and the sum runs over all nontarget items. This model has one parameter more than Eq. **S22**. The second modified mixture reflects the possibility that the nontarget weight depends on the distance (along the circle) between the target and the nontarget location,

$$p_{\text{fit}}(\hat{s}\,|\,s;w,\kappa_{\text{fit}}) = \frac{w_{\text{guess}}}{2\pi} + \frac{1-w_{\text{guess}}}{2\pi I_0(\kappa_{\text{fit}})}\frac{\sum_{j=1}^{N} w_{d_j} e^{\kappa_{\text{fit}}\cos(\hat{s}-s_j)}}{\sum_{j=1}^{N} w_{d_j}},$$

where d_j is the distance along the circle between the target and the jth item in units of the minimum distance. It can take integer values from 0 to 4, with 0 corresponding to the target. The normalization of the weights in the second term is needed because items occupy different sets of locations on different trials, but the overall distribution must always be normalized; therefore, the weights can only be relative. This mixture model has a total of six free parameters.

We compared the original descriptive mixture fit, Eq. **S22**, to its two variations. We applied the Bayesian information criterion to correct for the number of free parameters. When $\log L_{\max}$ is the maximum log likelihood of a model, the Bayesian information criterion (9) is $\text{BIC} = \log L_{\max} - \frac{k}{2}\log N_{\text{trials}}$, where k is the number of free parameters (two, three, or six) and N_{trials} is the number of trials.

Bayesian 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 (10). We explain the method for delayed estimation; for change localization, it is analogous. Each model m produces a predicted error distribution $p(\Delta s;m,N,\mathbf{t})$, where $\mathbf{t}$ denotes the model parameters. 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,

$$L(m) = p(\text{data}\,|\,m) = \int p(\text{data}\,|\,m,\mathbf{t})\,p(\mathbf{t}\,|\,m)\,d\mathbf{t} = \int\left(\prod_{i=1}^{N_{\text{trials}}} p(\Delta s_i;m,N_i,\mathbf{t})\right)p(\mathbf{t}\,|\,m)\,d\mathbf{t},$$

where Δs_i and N_i are the estimation error and set size on the ith trial, respectively. It is convenient to take the logarithm and rewrite it as

$$\log L(m) = \log L_{\max}(m) + \log\int \exp(\log L(m;\mathbf{t}) - \log L_{\max}(m))\,p(\mathbf{t}\,|\,m)\,d\mathbf{t}, \quad \textbf{[S23]}$$

where $\log L(m;\mathbf{t}) = \sum_{i=1}^{N_{\text{trials}}} \log p(\Delta s_i;m,N_i,\mathbf{t})$ and $L_{\max}(m) = \max_{\mathbf{t}} L(m;\mathbf{t})$. This form prevents numerical problems, because the exponential in the integrand of Eq. **S23** is now of order 1 near the maximum-likelihood value of $\mathbf{t}$. For the prior, we assume a uniform distribution across a plausible range (Table S1), whose size we denote S_j for the jth parameter. Then Eq. **S23** becomes

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

where dim $\mathbf{t}$ is the number of parameters. We approximated the integral through a Riemann sum, with 25 bins in each parameter dimension (we verified that this is a sufficiently large number to give robust results, Fig. S11*B*). 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 $\text{BIC} = \log L_{\max} - \frac{\dim\mathbf{t}}{2}\log N_{\text{trials}}$.

Numerical robustness. Because the model predictions for the VP models could not be computed analytically, we used Monte Carlo simulations. To obtain the model predictions, we drew 250 samples per combination of parameters, per stimulus. To verify whether 250 is a sufficiently high number, we checked how many samples are approximately needed for the model likelihoods to converge. The result shows that about 10 samples are needed (Fig. S11*A*). Hence, 250 samples is a sufficiently high number for obtaining reliable results.

For each model, we numerically approximated the integral over parameter space in the Bayesian model comparison by a Riemann sum. For all models, we discretized the parameter dimensions into 25 bins (except for K in the IL and SA models, because that parameter takes integer values only between 1 and 8). Results from running the analysis with different numbers of bins shows that about 15 bins are needed for convergence (Fig. S11*B*). Hence, 25 is a sufficiently high number of bins to obtain reliable results.

SI Results

Raw Data. An example of the descriptive mixture fits (*Methods*) to the histograms of estimation error of a single subject at all set sizes is shown in Fig. S2.

Scrolling vs. Color Wheel. Two conditions were used in experiment 1: responding using a color wheel and responding using scrolling. In the color wheel condition, subjects responded by a mouse click on an annulus composed of all 180 colors that were used for the stimulus array and centered at the center of the screen with a radius of 8.2° and a width of 2.2° (Fig. S3*A*). The color wheel was randomly rotated on each trial. In the scrolling condition, subjects used the left and right arrow keys to scroll through all possible colors. After the first key press, a random color appeared within the thicker circle. We find that w declines with set size, N, in both conditions (Fig. S3*B*). Using a two-way repeated-measures ANOVA with factors set size (1–8) and response modality (color wheel or scrolling), we find that w is significantly different between response modalities [main effect of set size, $F(2, 24) = 64.3$, $P < 0.001$; main effect of response modality, $F(1, 12) = 22.5$, $P < 0.001$]. At all set sizes except 1 and 8, a paired t test shows a significant difference between response modalities ($P < 0.01$). Estimated capacities are 4.5 ± 0.3 and 3.5 ± 0.3 in scrolling and color wheel conditions, respectively, constituting a significant difference [two-tailed paired t test, $t(12) = -2.94$, $P < 0.05$]. This result shows that the rate of (apparent) guessing increases with set size but is substantially smaller in the scrolling condition than in the color wheel condition.

Nontarget Reports. We investigated whether there was evidence for "nontarget responses" (i.e., responses in which subjects reported the color of a nontarget item). We compared the goodness-of-fit of the standard mixture model consisting of a uniform and a Von Mises component centered at the color value of the target with that of a model that also contained Von Mises distributions centered at the nontarget items. We found evidence for nontarget responses in the data from the color wheel condition but not in the data from the scrolling condition (Fig. S3*C*). Therefore, the scrolling condition was used for further analysis.

Parameter Estimates. Maximum-likelihood estimates of the parameters in all models in all experiments are shown in Table S1.

1. Palmer J (1990) Attentional limits on the perception and memory of visual information. *J Exp Psychol Hum Percept Perform* 16:332–350.
2. Shaw ML (1980) Identifying attentional and decision-making components in information processing. *Attention and Performance*, ed Nickerson RS (Erlbaum, Hillsdale, NJ), Vol VIII, pp 277–296.
3. Cover TM, Thomas JA (1991) *Elements of Information Theory* (Wiley, New York).
4. Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9:1432–1438.
5. Abramowitz M, Stegun IA, eds (1972) *Handbook of Mathematical Functions* (Dover, New York) .
6. Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. *Nature* 453:233–235.
7. Mardia KV, Jupp PE (1999) *Directional Statistics* (Wiley, New York).
8. Bays PM, Catalao RFG, Husain M (2009) The precision of visual working memory is set by allocation of a shared resource. *J Vis* 9:7–, 1–11.
9. Schwartz GE (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464.
10. MacKay DJ (2003) *Information Theory, Inference, and Learning Algorithms* (Cambridge Univ Press, Cambridge, UK).

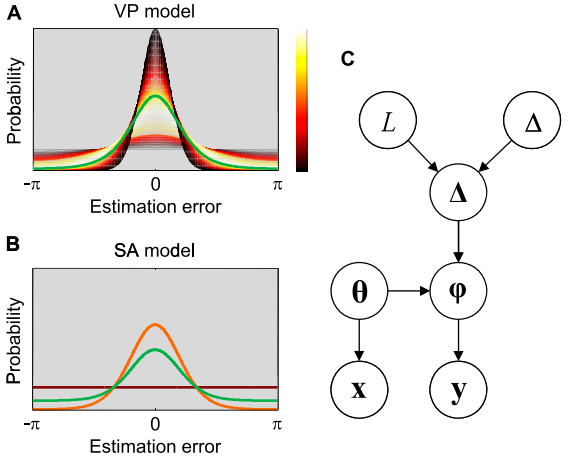

Fig. S1. (*A*) In the VP model, the error distribution in delayed estimation (green) is a mixture of a continuum of von Mises distributions with different *J* (color bar colors). Whiter colors represent a higher proportion in the mixture, according to a gamma distribution. (*B*) By contrast, in the SA model (here with $N = 5$ and $K = 3$), the error distribution (green) is a mixture of a uniform (red) and a Von Mises distribution (orange). (*C*) Generative model for the change localization task. *L*, location of the change; Δ, magnitude of change; $\boldsymbol{\Delta}$, 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.

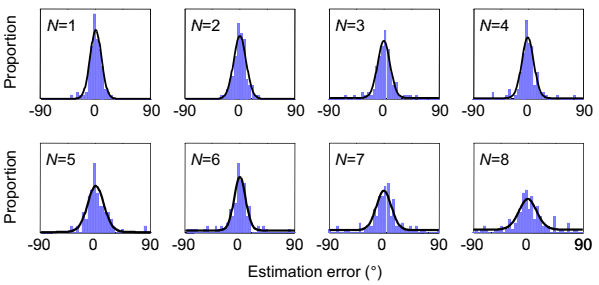

Fig. S2. Error distributions at all set sizes for subject 11 in experiment 1. Solid lines are the best fits of a mixture of a Von Mises and a uniform distribution. Note the systematic discrepancy, which is predicted by the VP model (Fig. 5 *B* and *E*).

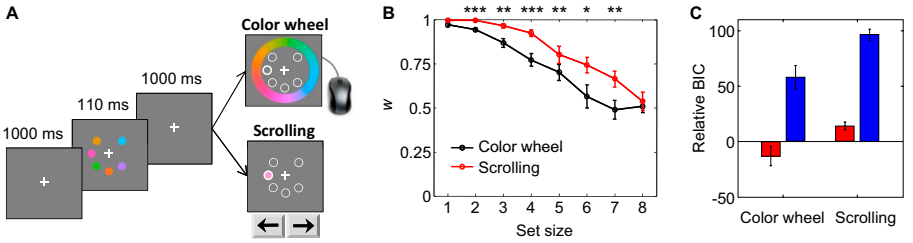

Fig. S3. Comparing response modalities (color wheel and scrolling) in delayed estimation of color. (*A*) Course of a trial. Two response modalities were used within subjects: clicking on a color wheel and scrolling through all possible colors. (*B*) Weight *w* (mean and SEM across subjects) in the scrolling and color wheel conditions. Significant differences: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. (*C*) Bayesian information criterion (BIC) for the model in which the error distribution is described as a mixture distribution without nontarget reports relative to the model containing nontarget reports, with the probability of reporting a nontarget being independent of (red) or dependent on distance (blue). Positive values indicate that it is more likely that no nontarget colors were reported. In the color wheel condition, but not in the scrolling condition, there is evidence of nontarget reports.

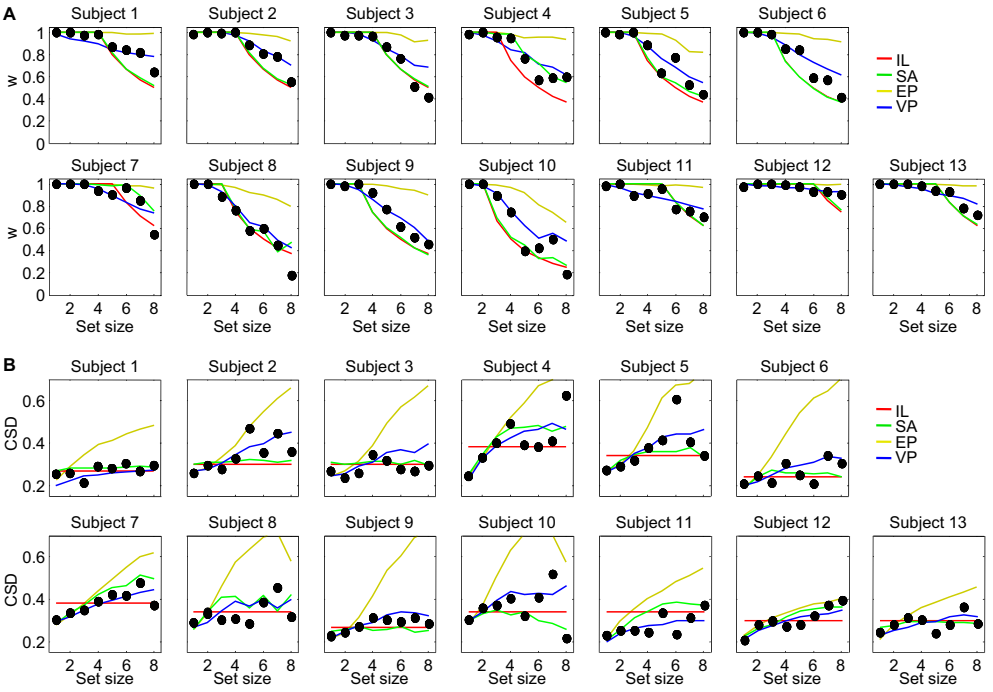

Fig. S4. Data and model predictions for the parameters *w* and CSD in individual subjects in experiment 1. (*A*) Weight *w*. (*B*) CSD. The estimates of *w* in the IL and SA models almost completely overlap for most subjects.

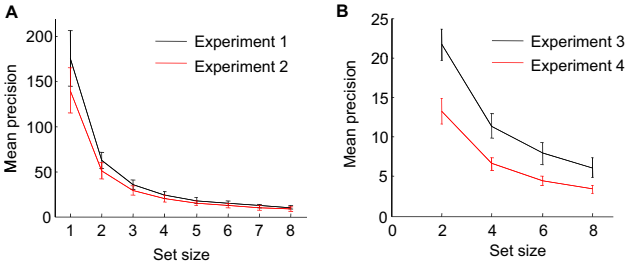

Fig. S5. (*A* and *B*) Estimated mean precision in the VP model as a function of set size, in delayed estimation (*A*) and change localization (*B*). The difference in scale might be due to observers deriving a greater performance benefit from increased attention in delayed estimation than in change localization.

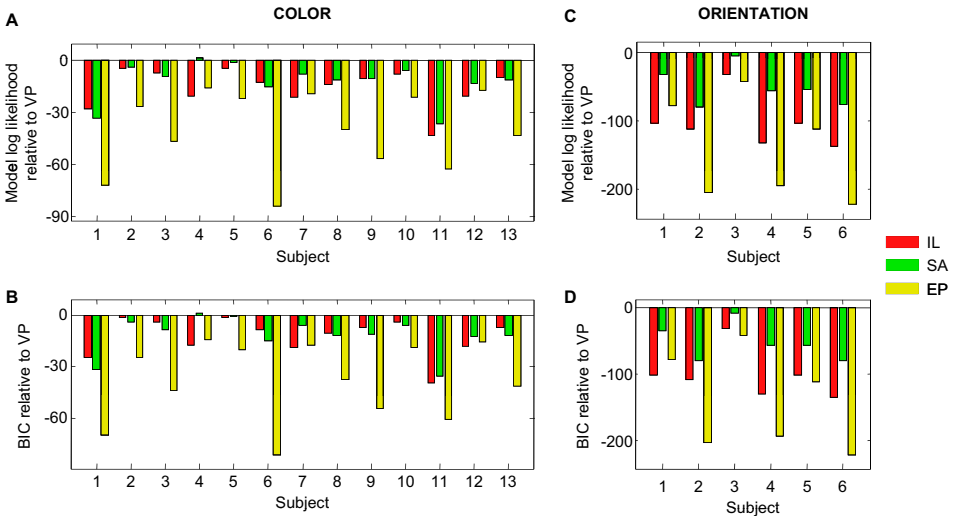

Fig. S6. Model comparison for individual subjects in delayed estimation. (*A*) Experiment 1, Bayesian model comparison. (*B*) Experiment 1, Bayesian information criterion. (*C*) Experiment 2, Bayesian model comparison. (*D*) Experiment 2, Bayesian information criterion.

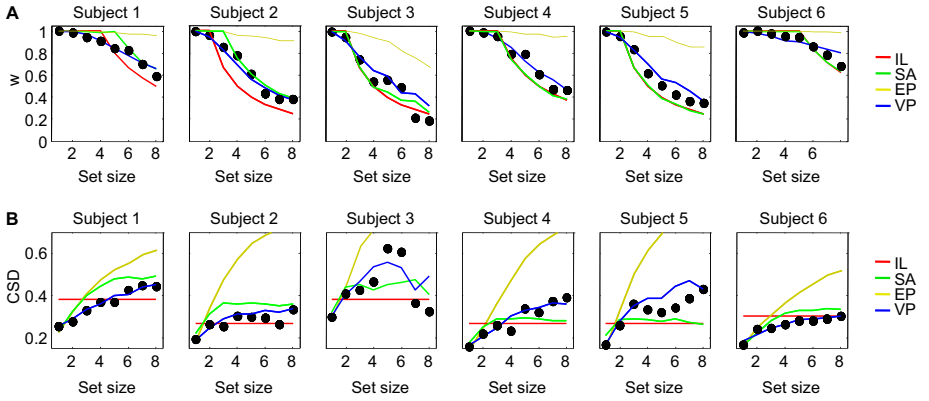

Fig. S7. Data and model predictions for the parameters *w* and CSD in individual subjects in experiment 2. (*A*) Weight *w*. (*B*) CSD. The predictions of *w* in the IL and SA models overlap for some subjects.

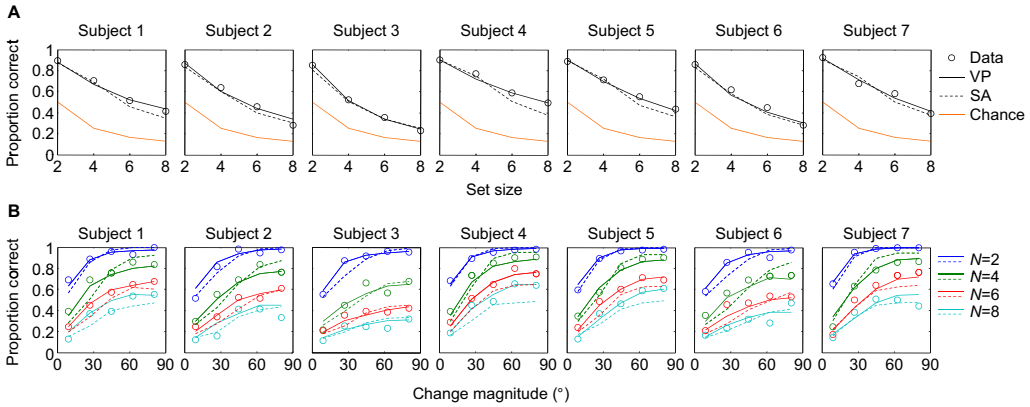

Fig. S8. Individual-subject fits of the SA and VP models in experiment 3 (solid line, VP; dashed line, SA; other models are not shown to avoid clutter). (*A*) Proportion correct as a function of set size. (*B*) Proportion correct as a function of change magnitude at each set size.

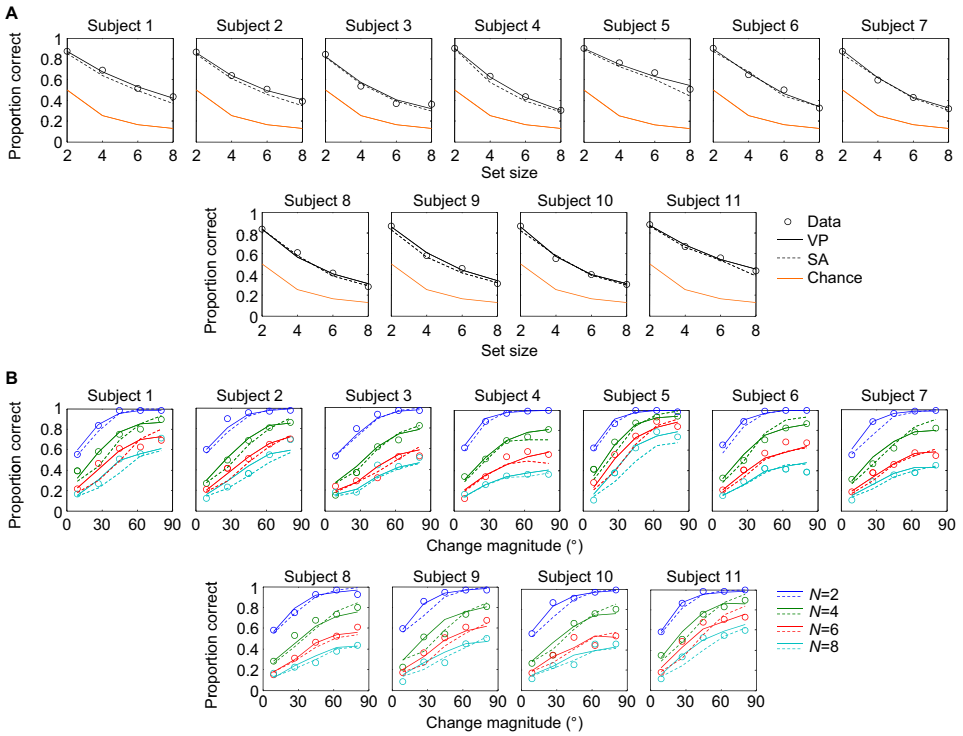

Fig. S9. Individual-subject fits of the SA and VP models in experiment 4 (solid line, VP; dashed line, SA; other models are not shown to avoid clutter). (*A*) Proportion correct as a function of set size. (*B*) Proportion correct as a function of change magnitude at each set size.

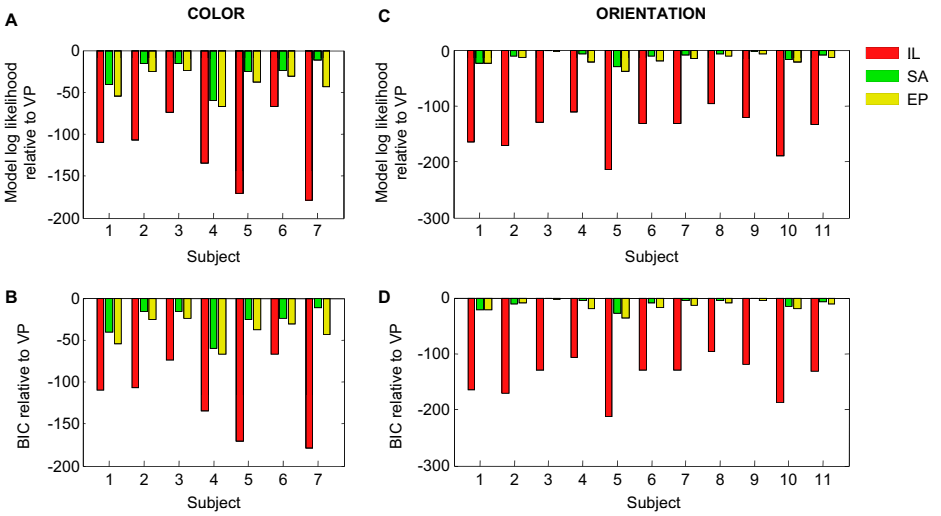

Fig. S10. Model comparison results for individual subjects in delayed estimation. (*A*) Experiment 3, Bayesian model comparison. (*B*) Experiment 3, Bayesian information criterion. (*C*) Experiment 4, Bayesian model comparison. (*D*) Experiment 4, Bayesian information criterion.

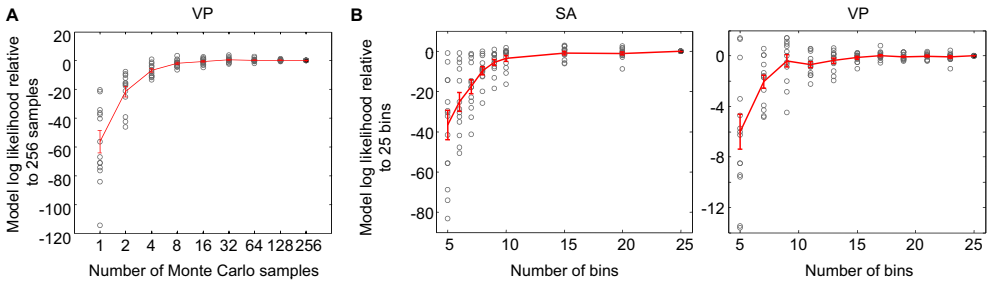

Fig. S11. Robustness of numerical results. (*A*) Effect of the number of Monte Carlo samples used to compute the VP model predictions on the model log likelihood in experiment 1. For each subject, we computed the model log likelihood for different numbers of samples and plotted it relative to the subject's model log likelihood when using 256 samples. Mean and 1 SEM are shown in red. (*B*) Effect of the number of bins used to approximate the integral over parameter space in computing the model log likelihood of the SA and VP models. Here, 250 samples were used to compute the model predictions.

Table S1. Means and SEs of the maximum-likelihood estimates and tested ranges of model parameters

	Delayed estimation							Change localization						
		Experiment 1		Experiment 2		Tested range			Experiment 3		Experiment 4		Tested range	
Model	Parameter	Mean	SEM	Mean	SEM	Min	Max	Parameter	Mean	SEM	Mean	SEM	Min	Max
IL	K	3.85	0.32	3.00	0.52	1	8	K	2.71	0.36	2.55	0.28	1	8
	κ_r	5.64	0.43	5.88	0.71	1	200	ε	0.279	0.035	0.282	0.021	0	1
SA	K	4.00	0.34	3.33	0.56	1	8	K	2.86	0.14	4.09	0.39	1	8
	J_1	18.5	5.2	4.04	0.72	1	50	J_1	5.9	1.0	3.51	0.72	1	15
	κ_r	29	15	193.4	6.6	1	200							
EP	J_1	106	32	22.5	5.2	10	300	J_1	11.2	1.8	11.1	1.6	1	25
	α	2.12	0.27	1.94	0.25	0	5	α	1.018	0.067	0.984	0.071	0	2
	κ_r	68	25	121	36	1	200							
VP	$\overline{J_1}$	176	30	86	25	10	600	$\overline{J_1}$	42.3	3.3	27.4	4.5	1	60
	α	1.33	0.14	1.41	0.15	0	2.5	α	0.974	0.090	0.993	0.075	0	2
	τ	43.5	7.2	27.9	8.8	1	100	τ	36.0	5.7	13.1	1.9	1	60
	κ_r	14.2	2.5	25.3	8.7	1	200							