Behavioral/Cognitive

Noise in Neural Populations Accounts for Errors in Working Memory

Paul M. Bays[1,2]

[1]Institute of Neurology, University College London, London, WC1N 3BG, United Kingdom, and [2]Institute of Cognitive and Brain Sciences, University of California, Berkeley, Berkeley, California 94720

Errors in short-term memory increase with the quantity of information stored, limiting the complexity of cognition and behavior. In visual memory, attempts to account for errors in terms of allocation of a limited pool of working memory resources have met with some success, but the biological basis for this cognitive architecture is unclear. An alternative perspective attributes recall errors to noise in tuned populations of neurons that encode stimulus features in spiking activity. I show that errors associated with decreasing signal strength in probabilistically spiking neurons reproduce the pattern of failures in human recall under increasing memory load. In particular, deviations from the normal distribution that are characteristic of working memory errors and have been attributed previously to guesses or variability in precision are shown to arise as a natural consequence of decoding populations of tuned neurons. Observers possess fine control over memory representations and prioritize accurate storage of behaviorally relevant information, at a cost to lower priority stimuli. I show that changing the input drive to neurons encoding a prioritized stimulus biases population activity in a manner that reproduces this empirical tradeoff in memory precision. In a task in which predictive cues indicate stimuli most probable for test, human observers use the cues in an optimal manner to maximize performance, within the constraints imposed by neural noise.

Key words: **divisive normalization; neural gain; Poisson noise; population coding; resource; short term memory**

Introduction

Short-term recall of basic visual features, such as orientation, declines steadily in precision as the number of features held in memory increases (Palmer, 1990; Wilken and Ma, 2004; Bays and Husain, 2008). Selection processes counteract this limitation by enhancing storage fidelity of behaviorally important visual items, but this incurs a cost to the precision of other items in memory (Bays et al., 2011a; Gorgoraptis et al., 2011; Melcher and Piazza, 2011; Lara and Wallis, 2012). These observations have lent support to cognitive models of working memory that invoke a limited supply of a representational medium, shared out between visual items (Bays and Husain, 2008; Zhang and Luck, 2008; Bays et al., 2009).

The pattern of recall errors in the space of possible feature values has proved an important testing ground for competing models of working memory. Observers make recall errors that consistently deviate from the familiar normal distribution. This has been interpreted as evidence for abrupt failures of recall, consistent with discrete or quantized memory resources (Zhang and Luck, 2008; Anderson et al., 2011). Alternatively, the same pattern of errors can be reproduced by continuous-resource models that assume random variability in resource allocation (Fougnie et al., 2012; van den Berg et al., 2012).

These studies have all assumed that errors in working memory representations are normally distributed and sought to identify a cognitive architecture that can account, on this basis, for the non-normal error distributions observed in recall tasks. An alternative approach is to consider the biological basis for memory representations in the nervous system and investigate how these representations fail.

In cortex, basic sensory parameters are encoded in the population activity of ensembles of neurons, each tuned to fire maximally in response to a preferred feature value. Errors occur in decoding because the processes underlying generation of neuronal spikes are probabilistic (Pouget et al., 2000). In the limit of large population activity, an efficient estimator will reconstruct a stimulus with errors that are normally distributed (Seung and Sompolinsky, 1993). However, decoding accuracy may deviate from asymptotic limits when asymptotic assumptions are not met (Xie, 2002; Berens et al., 2011). In the case of working memory, the substantial variability observed in short-term recall compared with sensory estimation suggests a neural signal that is weak compared with noise.

Here I examine errors produced by an ideal observer decoding stimulus parameters from the activity of probabilistically spiking neurons. As the gain of the neural population decreases, error distributions deviate from normality in the manner characteristic

Received July 26, 2013; revised Jan. 20, 2014; accepted Jan. 25, 2014.

Author contributions: P.M.B. designed research; P.M.B. performed research; P.M.B. contributed unpublished reagents/analytic tools; P.M.B. analyzed data; P.M.B. wrote the paper.

This research was supported by the Wellcome Trust. I thank Peter Dayan, Masud Husain, Wei Ji Ma, Loic Matthey, and Ronald van den Berg for helpful discussion and comments on this manuscript and Louise Marshall for technical assistance and additional data collection.

This article is freely available online through the *J Neurosci* Author Open Choice option.

Correspondence should be addressed to Dr. Paul Bays, 3210 Tolman Hall, University of California, Berkeley, Berkeley, CA 94720-1650. E-mail: p.bays@ucl.ac.uk.

DOI:10.1523/JNEUROSCI.3204-13.2014

of human working memory. For a neural population encoding more than one stimulus, a model in which total activity is held constant reproduces observed errors and has a plausible basis in divisive normalization (Carandini and Heeger, 1994, 2012).

Differential weighting of inputs to a normalized population can account for the tradeoff in memory precision between stimuli. In recall of multiple items with unequal test probability, I show that human performance matches predictions of an ideal observer that optimally weights neural activity to minimize error on the task.

Materials and Methods

Experimental methods. A total of 15 subjects (six males, nine females; aged 18–40 years) participated in the study after giving informed consent, in accordance with the Declaration of Helsinki. All subjects reported normal color vision and had normal or corrected-to-normal visual acuity. Stimuli were presented on a 21-inch CRT monitor with a refresh rate of 130 Hz. Subjects sat with their head supported by a forehead and chin rest and viewed the monitor at a distance of 60 cm. Eye position was monitored online at 1000 Hz using an infrared eye tracker (SR Research). Each subject took part in one of two experiments, described below.

Experiment 1: no cue. Each trial began with the presentation of a central white fixation cross (0.75° of visual angle) against a gray background. Once a stable fixation was recorded within 2° of the cross, a sample array consisting of one, two, four, or eight oriented colored bars (2° × 0.3°) was presented for 2 s. Bar orientations were independently chosen at random from the full range of possible orientations (0°–180°). Each bar was positioned on an invisible circle, radius 6°, centered on the fixation cross. Positions were chosen from a set of eight locations equally spaced around the circle. Colors were randomly selected on each trial, without repetition, from a set of eight highly discriminable colors.

After the sample array, the display was blanked for 1 s and then a probe display was presented, consisting of the reappearance of one randomly chosen bar from the sample display with a new randomly selected orientation. Participants used an input dial (PowerMate USB Multimedia Controller; Griffin Technology) to adjust the orientation of this probe bar to match the remembered orientation of the corresponding item in the preceding sample array (the target). Responses were not timed, and subjects were instructed to be as precise as possible. Any trial on which gaze deviated >2° from the central cross before the probe display was aborted and restarted with new feature values. Eight subjects participated in Experiment 1. Each subject completed a total of 900 trials, comprising 225 trials at each of the four different array sizes, randomly interleaved.

Experiment 2: informative cue. The procedure was identical to Experiment 1 except for the following modifications. On each trial, the sample array was immediately preceded by a cue stimulus, consisting of an annulus (white, radius 2.5°) presented for 500 ms at a location corresponding to one of the items in the upcoming sample array. Participants were informed that memory for the item appearing at the cued location was more likely to be tested than other items in the array. The cued item was chosen as the probe item with three times the frequency of each uncued item in the sample array. Seven subjects participated in Experiment 2. Each subject completed a total of 960 trials, comprising 320 trials at each of three different array sizes (two, four, or eight items), randomly interleaved. Trials were analyzed separately according to whether the probe corresponded to a cued or uncued item.

A relatively long (2 s) presentation duration was used in both experiments to ensure that errors reflected memory limits rather than incomplete sensory processing and because a previous study indicated that observers' ability to use predictive cues effectively may be diminished by brief exposures (Bays et al., 2011a).

Analyses. Orientations were analyzed and are reported with respect to the circular parameter space of possible feature values, i.e., the space of possible bar orientations [−90°, 90°) was mapped onto the circular space $[-\pi, \pi)$ radians. A measure of recall error was obtained on each trial in each experiment by calculating the deviation between the orientation reported by the subject and the correct (target) orientation. For each combination of experiment, subject, array size, and (in Experiment 2) cue validity, I calculated measures of variance and kurtosis for the observed recall errors, defining variance as the square of circular SD (σ^2) and using the definitions of SD and kurtosis for circular data given by Fisher (1995), as follows:

$$\sigma = \sqrt{-2 \log |\bar{m}_1|}$$

and

$$k = (|\bar{m}_2| \cos(\mathrm{Arg}(\bar{m}_2) - 2\,\mathrm{Arg}(\bar{m}_1)) - |\bar{m}_1|^4)(1 - |\bar{m}_1|)^{-2},$$

respectively, where $\bar{m}_n$ is the nth uncentered trigonometric moment.

Hypotheses regarding the effects of experimental parameters (array size, cue validity) on recall variance and kurtosis were tested by ANOVA and t tests.

Population coding model. I studied orientation encoding and decoding in a population of idealized neurons with spatial selectivity and orientation tuning (Pouget et al., 2000). Orientation information presented at each possible stimulus location provided feedforward input to an independent subpopulation of M neurons with bell-shaped tuning functions, such that the driving input to the ith neuron encoding a stimulus at the jth location was given by the following:

$$f_{ij}(\theta_j) = \exp(\omega^{-1}(\cos(\varphi_{ij} - \theta_j) - 1)), \quad (1)$$

where θ_j is the orientation of the stimulus at location j, φ_{ij} is the preferred orientation of the neuron, and ω determines the tuning curve width. Within each subpopulation, preferred directions were evenly distributed throughout the range of possible orientations. When no stimulus was presented at the location represented by a neuron, its driving input was 0.

Divisive normalization (Carandini and Heeger, 2012) operated over the entire population of neurons, such that the postnormalization output of a neuron (its firing rate) was determined by the normalization equation:

$$r_{ij}(\theta_j) = \gamma \frac{\alpha_j f_{ij}(\theta_j)}{\sum_{m,n} \alpha_n f_{mn}(\theta_n)}, \quad (2)$$

where γ is a gain constant (in hertz) that sets the overall responsiveness of the population, and α_j is a multiplicative gain factor related to attention at location j. Assuming that the distribution of tuning curves provides a dense uniform coverage of the orientation space (valid for large M), the summed activation of the population is independent of stimulus orientation, and Equation 2 simplifies to the following:

$$r_{ij}(\theta_j) = \gamma \frac{\alpha_j}{\sum_n \alpha_n} \frac{f_{ij}(\theta_j)}{M\bar{f}}, \quad (3)$$

where $\bar{f} = I_0(\omega^{-1})e^{-\omega^{-1}}$ is the mean activation of neurons responding to a stimulus (I_0 is the modified Bessel function of the first kind with order 0).

Persistent spiking activity was modeled as a homogeneous Poisson process, such that the probability of a neuron generating n spikes in time T is as follows:

$$\Pr[n_{ij}|\theta_j, T] = \frac{(r_{ij}(\theta_j)T)^{n_{ij}}}{n_{ij}!} \exp(-r_{ij}(\theta_j)T). \quad (4)$$

Recalling the orientation of a stimulus presented at a probed location p was modeled as maximum likelihood (ML) decoding of the population spiking activity $\mathbf{n}$ observed during a decoding interval T_d:

$$\hat{\theta}_p = \arg\max_{\theta_p} \Pr[\mathbf{n}|\theta_p, T_d] \quad (5)$$

$$= \arg\max_{\theta_p} \left(\sum_i^M n_{ip} \log\left(r_{ip}(\theta_p)\right) - \sum_i^M r_{ip}(\theta_p) T_d \right). \quad (6)$$

If two or more orientations tied for the ML, the decoded orientation was sampled at random from the tied values. Assuming dense uniform coverage, the second term is constant and Equation 6 further simplifies to the following:

$$\hat{\theta}_p = \arg\max_{\theta_p} \sum_i^M n_{ip} \log\left(f_{ip}\left(\theta_p\right)\right) \quad (7)$$

$$= \arg\max_{\theta_p} \sum_i^M n_{ip} \cos\left(\varphi_{ip} - \theta_p\right). \quad (8)$$

The decoded orientation has some error relative to the original stimulus value: $\Delta\theta = \hat{\theta}_p \ominus \theta_p$ (where $\ominus$ indicates subtraction on the circle). The distribution of this error is uniquely determined by the values of two terms: the tuning width ω, and a second term, ξ, which can be interpreted as the expected summed activity during the decoding window of neurons encoding the probed stimulus. To see why, it is convenient to consider the combined activity of neurons selective for the probed location in terms of the preferred stimulus corresponding to each spike: $\{\varphi_{(1)}, \varphi_{(2)}, \ldots \varphi_{(m)}\}$, where the notation $\varphi_{(1)}$ indicates the preferred stimulus value of the neuron that generated the ith of m spikes. The decoded orientation (Eq. 8) can then be written as follows:

$$\hat{\theta}_p = \arg\max_{\theta_p} \sum_i^m \cos\left(\theta_p - \varphi_{(i)}\right). \quad (9)$$

Because spikes are generated by independent Poisson processes, every spike event is conditionally independent of every other given the true stimulus value, i.e., $p(\varphi_{(i)} \mid \theta_p, \varphi_{(j)}) = p(\varphi_{(i)} \mid \theta_p)\ \forall\, j \neq i$. Approximating the uniformly spaced discrete distribution of preferred directions of M neurons by a continuous uniform distribution, this probability is given by the following:

$$p(\varphi_{(i)}|\theta_p) = \frac{\exp(\omega^{-1}\cos(\varphi_{(i)} - \theta_p))}{2\pi I_0(\omega^{-1})}. \quad (10)$$

The distribution of m, the total spike count during the decoding window, being a sum of M independent Poisson distributions (Eq. 4), is itself Poisson with expected value $\xi = \gamma T_d \frac{\alpha_p}{\sum_n \alpha_n}$. Defining $\varepsilon_{(i)} = \varphi_{(i)} \ominus \theta_p$ leads to a very concise description of the distribution of decoding error $\Delta\theta$ in terms only of ω and ξ:

$$\Delta\theta = \arg\max_{\theta} \sum_i^m \cos\left(\theta - \varepsilon_{(i)}\right), \quad (11)$$

$$m \sim \text{Poisson}(\xi), \quad (12)$$

$$p(\varepsilon_{(i)}) = \frac{\exp\left(\omega^{-1}\cos\varepsilon_{(i)}\right)}{2\pi I_0\left(\omega^{-1}\right)}. \quad (13)$$

Simulation and model fitting. I used Monte Carlo simulation to investigate the distribution of error under the model for different values of ω and ξ and to compare predictions of the model with the empirical distribution of errors obtained in Experiments 1 and 2.

I generated 10^6 samples from the error distribution $p(\Delta\theta)$ for every element in a logarithmically spaced 50×50 element grid of values of ω (range of 2^{-4} to 2^2) and ξ (range of 2^0 to 2^6), following Equations 11–13, where the maximization in Equation 11 was based on values calculated at 10^3 evenly spaced points in the interval $[-\pi, \pi)$. From these samples, I calculated a histogram estimate of the error distribution corresponding to each grid point, based on 50 equally spaced bins (I considered other levels of discretization: likelihoods converged in the range of 25–50 bins, and additional increases beyond this number of bins had no effect on results).

To compare the output of the model to results from Experiment 1, in which all stimuli were equally likely to be probed, I set the attentional gain factor $\alpha = 1$ for all locations where stimuli were presented (setting $\alpha = 0$ for empty locations), i.e., all stimuli were weighted equally. I set the decoding interval $T_d = 100$ ms. This left two free parameters: ω, the tuning width, and γ, the gain constant that determines the total output activity of the population. These parameters were fitted to the data obtained from each subject separately, based on maximizing their likelihood given the observed errors: $\mathcal{L}(\omega, \gamma \mid \Delta\theta, \mathbf{N})$.

To capture performance in Experiment 2, in which a predictive cue preceded each stimulus array, I modified the attentional gain factor corresponding to the cued location. Specifically, the gain factor corresponding to the cued item, α_{cued}, was treated as a free parameter with a separate value for each array size, whereas gain factors for uncued stimulus locations were held constant, $\alpha_{\text{uncued}} = 1$. (As above, $\alpha = 0$ for locations without a stimulus.) ML estimates for the five parameters (ω, γ, $\alpha_{\text{cued}}^{N=2}$, $\alpha_{\text{cued}}^{N=4}$, $\alpha_{\text{cued}}^{N=8}$) were obtained for each subject's data separately using the same procedure as above.

For each subject in Experiments 1 and 2 and each array size, I calculated optimal values of the gain factor α_{cued} that would minimize the variance of error over a sequence of trials, given the relative frequency with which cued and uncued items were probed in Experiment 2 and the values of ω and γ obtained by ML. Specifically, I calculated the variance for 100 logarithmically spaced values of α_{cued} in the range of 0.01–100, based on 10^6 samples drawn from the error distribution under the model, sampling cued and uncued locations in the same proportions as in the actual experiment. The estimate of optimal α_{cued} was the value that produced the smallest variance [note that identical results would have been obtained by minimizing an explicit cost function of the form $C = 1 - \cos(\Delta\theta)$].

Unprobed stimuli. To test whether the unprobed stimuli in memory influenced responses, I fit a variant of the model in which decoding of the correct neural subpopulation (the one corresponding to the probed location) was probabilistic. Specifically, I defined a parameter ϵ as the probability of incorrectly reporting the neural estimate corresponding to any individual unprobed stimulus. Likelihoods under the model were calculated for 50 values of ϵ in the range of 0–0.14.

Baseline activity and signal-to-noise ratio. I examined the effect of adding baseline (background) activity to the model neurons by replacing the activation function (Eq. 1) with a modified function:

$$f_{ij}(\theta_j) = \exp(\omega^{-1}(\cos(\varphi_{ij} - \theta_j) - 1)) + f_{(0)}, \quad (14)$$

where $f_{(0)}$ was a parameter determining the baseline activation level. The assumption of dense uniform coverage remained valid in this model, and the decoding error could be shown to depend only on ω, ξ, and $f_{(0)}$ by a similar argument to the above. I approximated the likelihood using Monte Carlo simulation, generating 10^5 samples from the error distribution $p(\Delta\theta)$ for every element in a logarithmically spaced $50 \times 50 \times 50$ element grid of values of ω (range of 2^{-4} to 2^4), ξ (range of 2^{-2} to 2^{23}), and $f_{(0)}$ (range of 2^{-6} to 2^4). Figure 4 plots baseline activation as a proportion of peak activation $f_{(0)}/(f_{(0)} + 1)$.

The signal-to-noise ratio (SNR) per neuron is the ratio between variance attributable to stimulus differences in the expected spike count during the decoding window, $\bar{n}(\theta) = T_d r(\theta)$, and the additional variance in actual spike counts $n(\theta)$ attributable to Poisson noise:

$$\text{SNR} = \frac{\text{Var}_\theta[\bar{n}(\theta)]}{\text{Var}_\theta[n(\theta)|\bar{n}(\theta)]} = \frac{T_d{}^2 \text{E}_\theta[r(\theta)^2] - (T_d \text{E}_\theta[r(\theta)])^2}{T_d \text{E}_\theta[r(\theta)]}$$

$$= \frac{T_d \gamma}{MN}\left(\frac{I_0\left(2\omega^{-1}\right)}{I_0(\omega^{-1})^2} - 1\right), \quad (15)$$

based on equal attentional gain for all stimuli and no baseline activity. Incorporating a baseline activation $f_{(0)}$ as described above, I obtained the following more general formula:

$$\text{SNR} = \frac{T_d \gamma}{MN} \frac{I_0\left(2\omega^{-1}\right) - I_0\left(\omega^{-1}\right)^2}{\left(I_0\left(\omega^{-1}\right) + e^{\omega^{-1}} f_{(0)}\right)^2}. \quad (16)$$

Heterogeneous and non-normal tuning. To investigate the effects of variability in shape and amplitude of tuning functions on errors in the population coding model, I simulated ML decoding of populations of neurons with tuning functions of the general form (normal tuning),

$$f_{ij}(\theta_j) = a_{ij}\exp(\omega_{ij}^{-1}(\cos(\varphi_{ij} - \theta_j) - 1)) + f_{(0)ij} \quad (17)$$

and (cosine tuning),

$$f_{ij}(\theta_j) = \begin{cases} a_{ij}\cos\left(\omega_{ij}^{-1}(\varphi_{ij} \ominus \theta_j)\right) + f_{(0)ij} & \text{for} \quad \omega_{ij}^{-1}|\varphi_{ij} \ominus \theta_j| < \dfrac{\pi}{2} \\ f_{(0)ij} & \text{for} \quad \omega_{ij}^{-1}|\varphi_{ij} \ominus \theta_j| > \dfrac{\pi}{2} \end{cases}, \tag{18}$$

where amplitudes a_{ij}, tuning widths ω_{ij}, and baselines $f_{(0)ij}$ were normally distributed random variables (truncated at 0), and preferred orientation φ_{ij} was a uniformly distributed random variable on the circle. The decoder had knowledge of the individual tuning functions of the neurons. Error distributions were obtained by explicit simulation and decoding of spike trains generated by $M = 100$ or 1000 simulated neurons, following Equations 2–6.

Correlated activity. To investigate the effect of short-range noise correlations on decoding errors, I modified the (zero-baseline, homogeneous) population coding model to incorporate pairwise correlations in spike generation. The correlation in activity between the ith and jth neurons both responding to a stimulus at location k was an increasing function of the similarity in their preferred orientations:

$$c_{ik,jk} = c_0 \exp(-|\varphi_{ik} \ominus \varphi_{jk}|). \tag{19}$$

Activity of neurons selective to different locations was uncorrelated. Error distributions were again obtained by explicit simulation and decoding of spike trains generated by $M = 100$ or 1000 simulated neurons. The decoder did not have knowledge of the correlations: decoded values were obtained according to Equation 8. Correlated spike counts were generated based on the "latent Gaussian" method of Macke et al. (2008), using code available at http://bethgelab.org/software/mvd/.

Discrete- and continuous-representation models. I compared the population coding model to two previously proposed cognitive-level models based on distributing a limited representational medium between items.

In the discrete-representation (or "slots + averaging") model (Zhang and Luck, 2008), the representational medium is divided into a fixed number of memory quanta or "slots," M, which are distributed between array items. Each of the slots holds a separate representation of an object feature, equal to the correct feature value plus circular normally distributed noise with SD σ_1. Slots are distributed as evenly as possible between array items, and if there are more slots than array items, the same item may be represented more than once. For items that receive more than one slot, the multiple representations are averaged, such that recall error for an item with S slots is distributed as a circular normal with SD $\sigma_1/\sqrt{S}$.

On a trial with N items, each item receives one of two numbers of slots S_{high} or S_{low}, with probability P_{high} and P_{low}, respectively:

$$\begin{array}{ll} S_{\text{high}} = \lfloor M/N \rfloor + 1, & P_{\text{high}} = (M \bmod N)/N \\ S_{\text{low}} = \lfloor M/N \rfloor, & P_{\text{low}} = 1 - (M \bmod N)/N \end{array}, \tag{20}$$

where $\lfloor x \rfloor$ is the floor function. So the distribution of responses $\hat{\theta}$ for recall of a stimulus value θ is given by the following:

$$p(\hat{\theta}) = P_{\text{low}}\phi_{\text{VM}}\left(\hat{\theta};\, \theta,\, \sigma_1/\sqrt{S_{\text{low}}}\right) + P_{\text{high}}\phi_{\text{VM}}\left(\hat{\theta};\, \theta,\, \sigma_1/\sqrt{S_{\text{high}}}\right), \tag{21}$$

where $\phi_{\text{VM}}(\hat{\theta};\, \theta,\, \sigma)$ is the Von Mises probability density function with mean θ and circular SD σ. (The circular SD of a Von Mises distribution is related to its concentration κ by $\sigma = \sqrt{-2\log(I_1(\kappa)/I_0(\kappa))}$.)

For each subject in Experiment 1, I calculated likelihoods for a 50×50 element grid of parameters σ_1 (range of 0.3–30) and M (range of 1–50) of the discrete-representation model. To fit the data from Experiment 2, I introduced additional free parameters, $A^{N=2}$, $A^{N=4}$, and $A^{N=8}$ (each with range of 0 to M), that determined the number of slots allocated to the cued item at each set size; the remaining slots were allocated as evenly as possible between uncued items.

The variable-resource model (van den Berg et al., 2012) is a doubly stochastic model, in which items are represented with circular normal (Von Mises) distributed error, the variability of which is itself a random variable, resulting in an infinite mixture of circular normal distributions of different widths. van den Berg et al. parameterized the width of the Von Mises distribution by its Fisher information J, which is related to its concentration κ by $J = \kappa I_1(\kappa)/I_0(\kappa)$. According to the variable-resource model, the Fisher information has a gamma distribution:

$$p(J) = \frac{J^{\bar{J}/\tau - 1} e^{-J/\tau}}{\Gamma(\bar{J}/\tau)\tau^{\bar{J}/\tau}}, \tag{22}$$

where τ is a scale parameter that remains fixed across changes in array size N, and $\bar{J}$ is the mean of the distribution that has a power law relationship with array size: $\bar{J} = \bar{J}_1/N^{\alpha}$. For each subject in Experiment 1, I calculated likelihoods for a $50 \times 50 \times 50$ element grid of parameters $\bar{J}_1$ (range of 1–60), α (range of 0–2), and τ (range of 1–60) of the variable-resource model. Likelihoods for the variable-resource model were estimated by approximating the infinite mixture of normals by a discrete mixture consisting of 1000 normal distributions with concentrations κ logarithmically spaced in the range 10^{-3} to 10^{3} [this was found to be a more reliable, although asymptotically equivalent, method to the Monte Carlo approach used previously (van den Berg et al., 2012)].

Model comparison. Formal model comparisons were conducted based on Bayes factors, Bayesian information criterion (BIC) and Akaike information criterion (AIC). The Bayes factor is the ratio of likelihoods under two competing models, averaged across model parameters. Using Bayes factors automatically penalizes models with too much model structure, (i.e., too many free parameters). Bayes factors were calculated from the grid of likelihood values under each model by trapezoidal numerical integration, assuming a uniform prior over model parameters (for additional details of this method, see van den Berg et al., 2012). I also calculated the BIC and AIC, which are similar measures based on the maximum rather than average likelihood and incorporating explicit penalty terms for free parameters. A model recovery analysis (based on generating 2880 synthetic datasets using ML parameters obtained by fitting each of the models to data from Experiment 1) found all three measures to be similar in reliability, selecting the correct model on 93.6% (Bayes factor), 91.3% (BIC), and 91.5% (AIC) of datasets.

Results

I tested the ability of a population coding model to describe errors in short-term memory for displays consisting of multiple orientation stimuli. Memorization of the display was modeled as encoding of the stimulus orientations in the activity of a population of idealized memory neurons, with spatial selectivity and orientation tuning. Each orientation stimulus provided a feedforward input to a discrete subpopulation of neurons selective to its location. Within that subpopulation, the driving input to each neuron was determined by the deviation between the stimulus orientation and the preferred orientation of the cell, following a bell-shaped tuning function (Fig. 1*a*).

Divisive normalization operated over the population of neurons, such that the firing rate of each neuron was determined by dividing its individual driving input by the summed input to the population. As a consequence, the total response of the population was constant, and the summed firing rate of neurons encoding any individual stimulus was inversely related to the total number of stimuli represented.

Retention of the stimulus display in memory was modeled as persistent spiking activity in the population of neurons. Each neuron generated spikes stochastically and independently, according to a Poisson process, maintaining the firing rate determined by its normalized driving input (Fig. 1*b*). Subsequent recall of the orientation of a probed item was modeled as ML decoding of the population spike trains over a fixed time window. Because spike generation was stochastic, the decoded orientation was an imprecise estimate of the original stimulus value.

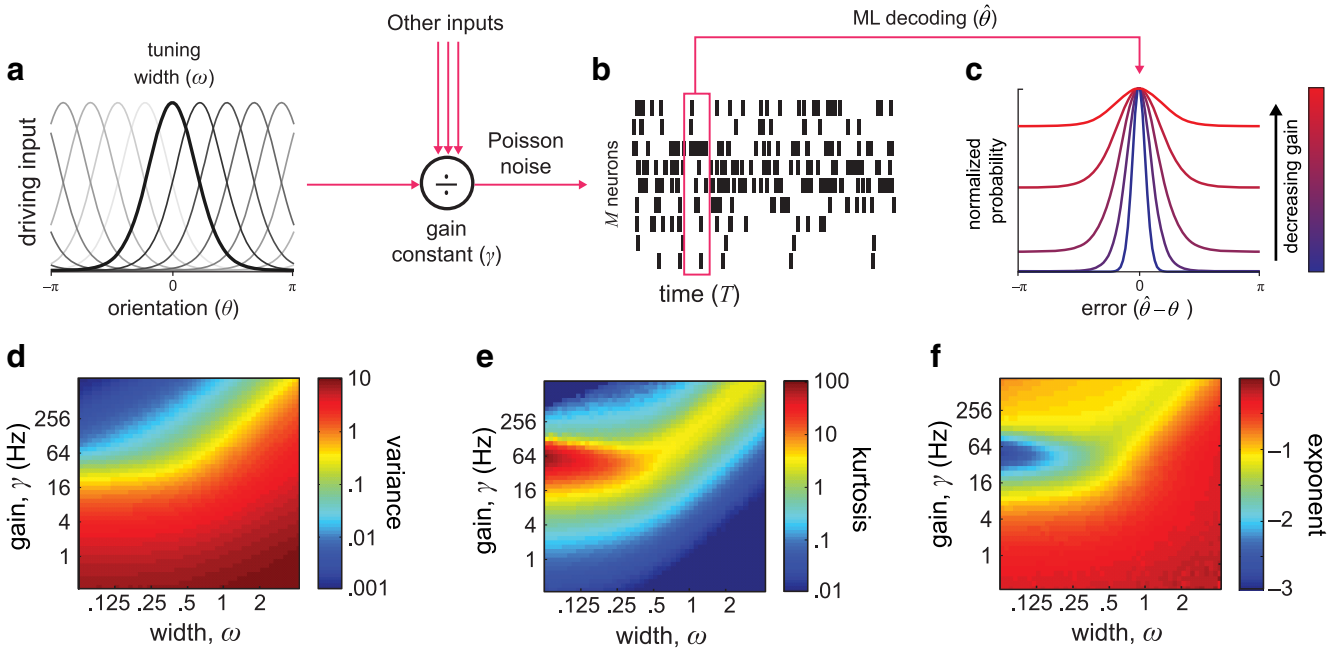

Figure 1. The population coding model. ***a***, Stimulus orientations were encoded in the activity of idealized neurons with preferred orientations evenly distributed on the circle and bell-shaped tuning functions (width ω). Each of N stimuli was encoded by an independent subpopulation of M neurons. Divisive normalization operated across the whole population, scaling population activity to a level determined by the gain constant, γ. (Note that the 180° range of orientations is represented here by the circular parameter space $-\pi$ to π). ***b***, Each neuron generated spikes according to a Poisson process, with mean firing rate determined by the normalized input of a neuron. Subsequent recall was modeled as ML decoding of the spiking activity of the subpopulation of neurons corresponding to a probed stimulus over a fixed time window. ***c***, Simulations showed that error in the recalled orientation depended on gain (summed output) of the decoded subpopulation, which declined with increasing N as a result of divisive normalization. At high gains, errors had an approximately normal distribution (e.g., blue curves). As gain decreased (magenta to red curves), variability increased and error distributions deviated from normality. Note that error distributions shown are normalized by peak probability to better illustrate distribution shape. ***d***, Variance of simulated errors under the population coding model as a function of gain and tuning width. The lowest variances (blue) were obtained with high gains and narrow tuning functions. The range of variances typically observed in human recall corresponds approximately to the yellow band. ***e***, Kurtosis of errors. Circular kurtosis approached 0 at the highest gains, indicating that errors had an approximately circular normal distribution. Kurtosis also approached 0 at the lowest gains, as errors approached a uniform distribution on the circle. Positive kurtosis (hot colors) was observed at intermediate gains for all tuning widths, indicating deviations from circular normality. ***f***, Exponent of a power law relating gain to error variance (estimated from the change of variance resulting from halving the gain, equivalent to doubling the number of items in memory). At high gains, the exponent approached −1, indicating variance was inversely proportional to gain (Seung and Sompolinsky, 1993). However, at intermediate gains, the exponent became more strongly negative (less than −1; green and blue regions) for all tuning widths.

Figure 1*c* provides an illustration of how the distribution of errors in a decoded stimulus parameter depended on the gain (summed firing rate) of the pool of neurons encoding it. For high gains, variability in the decoded parameter was small, and errors approximated a normal distribution centered on the true stimulus value (e.g., blue curve). As gain decreased, variability in the decoded parameter increased (e.g., magenta, red curves), but as well as becoming broader, the distribution of errors deviated systematically from the circular normal distribution.

Figure 1, *d* and *e*, plots variance and kurtosis (a measure of deviation from the normal distribution) of the decoded error distribution as a function of population gain and tuning width. Non-normal error distribution (positive kurtosis) was a universal observation for the population coding model at lower gains, although the gain at which the strongest deviations from normality were observed and the precise shape of the distribution depended on the width of the tuning function (see Fig. 4*a*,*b*). As gain fell to 0, decoding precision also approached 0, and errors approached a uniform distribution. Because the uniform distribution on the circle is also a circular normal, kurtosis fell to 0 at the lowest gains (Fig. 1*e*).

Population coding model accounts for human recall errors

To compare the predictions of the population coding model to human memory performance, observers in Experiment 1 were presented with displays consisting of between one and eight oriented bars. After a brief blank display, one location was probed, and subjects reported the remembered orientation of the item at that location using a response dial. Recall error was quantified as the angular deviation between the reported orientation and its true value in the memory array. The distributions of recall errors for stimulus arrays containing different numbers of orientations are plotted as black symbols in Figure 2*a* (for one representative subject) and Figure 2*c* (mean ± SE for eight subjects). As observed in previous studies, each increase in memory load resulted in a corresponding decrease in recall precision (broader error distributions, left to right).

I fit two global parameters of the population coding model, the tuning curve width ω and the gain constant γ, individually to each subject's data. Red lines in Figure 2*a* plot the distribution of errors generated by the model at each array size, with ML parameters for the representative subject ($\omega = 0.65$; $\gamma = 113$ Hz). As a result of divisive normalization, the activity encoding each stimulus orientation in the model decreased as array size increased, reproducing the increase in variability seen in empirical data. Figure 2*c* presents mean predictions of the fitted population coding model for the group (red lines; with ML parameters displayed in Fig. 2*b*; $\omega = 0.52 \pm 0.05$; $\gamma = 119 \pm 4$ Hz).

Figure 2*d* (black symbols) plots the variance (squared circular SD) of observers' responses as a function of array size. As observed in previous studies (Bays and Husain, 2008; Bays et al., 2009), the relationship between variance and load was not linear

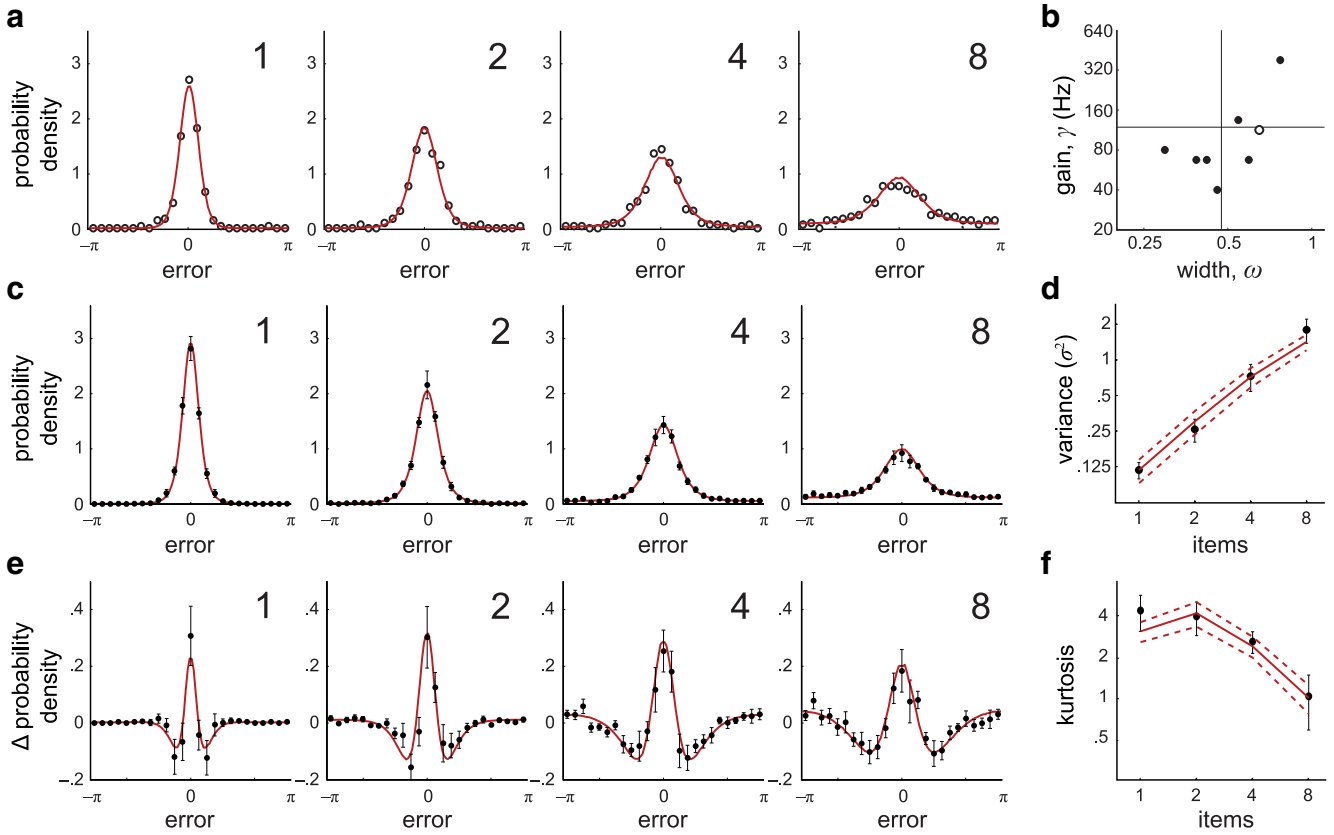

Figure 2. Human recall errors and model fits. ***a***, Black symbols show distribution of recall errors made by a representative human observer for displays of one to eight orientation stimuli. Red curves show error distributions generated by the population coding model for ML parameters γ and ω. Note that the model reproduced the changes in error distribution with array size (left to right) despite model parameters remaining fixed. ***b***, ML values of population gain γ and tuning width ω for eight subjects. Open circle corresponds to subject shown in ***a***. ***c***, Black symbols show mean distribution of recall errors for the group (error bars indicate ± 1 SE). Red curves show mean error distributions for the population coding model with ML parameters. ***d***, Variance of error in recalled orientation for human observers (black symbols) and for the population coding model with ML parameters (red curve, dashed lines indicate ± 1 SE). Variance has an approximately power-law relationship with set size (appearing linear on the log–log plot). ***e***, Deviation from the circular normal distribution. Black symbols plot mean discrepancy between subject error frequencies shown in ***c*** and circular normal (Von Mises) distributions matched in variance. Red curves plot equivalent deviations for the population coding model with ML parameters. ***f***, Kurtosis of recall errors for subjects (black symbols) and the population coding model (red curve, dashed lines indicate ± 1 SE). The circular normal distribution has kurtosis around 0.

but instead approximated a power law, appearing linear on the log–log axes in Figure 2*d* (exponent 1.36 ± 0.08, differing significantly from unity: $t_{(7)} = 4.5, p = 0.0027$). Results from the model (red line) closely approximated this nonlinear relationship. Such supralinear increases in variance were a feature of the model for a wide range of tuning widths and population gains (Fig. 1*f*).

Figure 2*e* (black symbols) plots the discrepancy between the observed distribution of errors made by observers at each array size and a circular normal distribution with the same variance. The distinctive "Mexican hat" function indicates excess kurtosis, i.e., that the error distributions have sharper peaks and longer tails than a normal distribution. Previous studies have noted the long tails of the distribution at high memory loads (Zhang and Luck, 2008). However, as this plot makes clear, such deviations from normality are not unique to high loads and are present even for one item (Fig. 2*e*, left panel). The circular kurtosis, a measure of deviation from the circular normal, is plotted in Figure 2*f* (black symbols). Kurtosis displayed a nonlinear relationship with set size, with the highest values obtained at lower memory loads.

As indicated previously (Fig. 1*c*), deviations from the normal distribution were characteristic of the population coding model at lower gains. Red lines in Figure 2*e* plot the deviation from normality of error distributions generated by the model with ML parameters. The model successfully replicated the discrepancies from normality observed in human performance, as well as the nonlinear relationship between kurtosis and memory load (Fig. 2*f*, red line; note that, for ML parameters, peak kurtosis occurred in the region of one to two items, accounting for the decline in kurtosis at larger set sizes).

Weighting of population activity prioritizes cued items

Previous studies showed that observers can maintain goal-relevant visual information with enhanced precision, at a cost to recall of other visual items (Bays and Husain, 2008; Gorgoraptis et al., 2011). In the population coding model, such a precision tradeoff could arise from an increase in input drive to neurons encoding the prioritized stimulus. As a result of divisive normalization, this would have the effect of both increasing recall fidelity for the preferred stimulus and reducing precision for other items in memory compared with the situation with equal inputs.

I performed a second experiment to test this account. The task was identical to Experiment 1, except that one item on each trial was highlighted by a visual cue appearing before the array, and this cued item was more likely to be probed on the subsequent test than other items in the array. Based on previous results, I expected that observers would prioritize accurate storage of the cued item to maximize their overall response precision.

The distributions of errors in Experiment 2, for stimulus arrays containing between two and eight items, are plotted as black symbols in Figure 3*a*: top row shows the overall distribution of

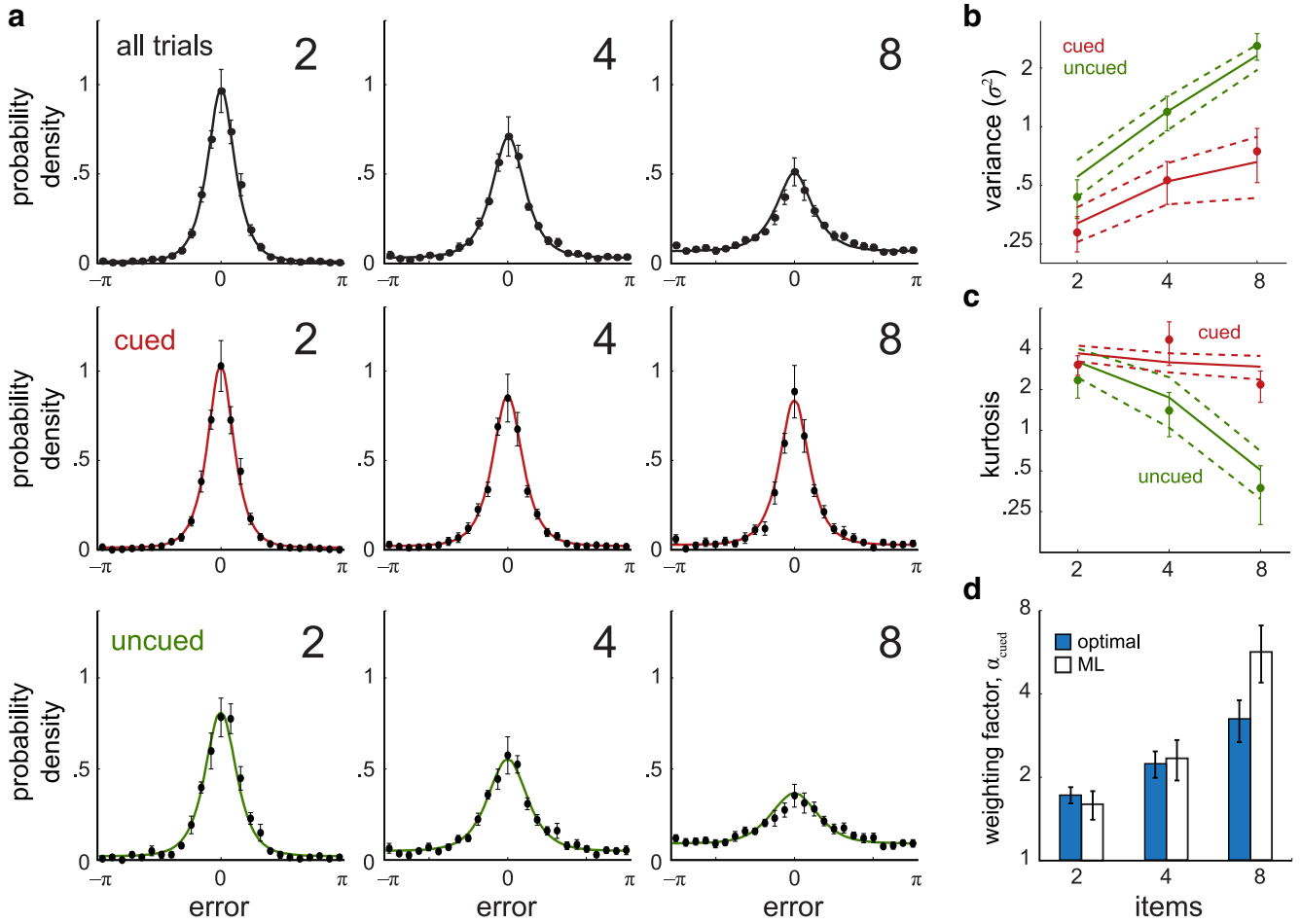

Figure 3. Recall with an informative cue. ***a***, Black symbols show mean distribution of recall errors for displays of two to eight orientations in Experiment 2, in which one stimulus, indicated by a cue, was more likely to be selected for test. Overall error variability increased with array size (top, left to right). When separating trials according to cue validity, recall was found to be consistently less variable for the cued item (middle) than for uncued items (bottom). In the population coding model, this corresponds to an increased weighting of activity related to the cued stimulus. Colored curves show mean error distributions generated by the model with ML parameters. Note that only the weighting of activity to stimuli differed across array sizes (left to right), with total population activity and tuning width remaining constant for each subject. ***b***, Variance of error for cued items (red) and uncued items (green), for human observers (symbols) and the population coding model with ML parameters (solid lines; dashed lines indicate ±1 SE). Differences in variance between cued and uncued items are accounted for in the model by differential weighting of activity. ***c***, Kurtosis of errors for cued and uncued items. ***d***, Comparison of weighting factors estimated from subject errors on the cued task (Experiment 2, white bars) and optimal weightings that would be expected to minimize total error variance (blue bars, based on ML parameters γ and ω obtained from the uncued task, Experiment 1). Optimal behavior predicts that weighting of activity in favor of the cued item should increase with each increase in display size (blue bars, left to right), and this predicted pattern was observed in the weighting factors estimated from subject errors (white bars).

recall errors, middle row shows only trials in which memory for the cued item was tested, and bottom row shows trials in which one of the uncued items was tested. Consistent with previous studies, cued items were recalled with significantly lower variability than uncued items (Fig. 3*b*, red symbols, cued; green symbols, uncued; $F_{(1,6)} = 15.4$, $p = 0.008$).

Within the population coding model, weighting of storage precision was modeled by a parameter, α_{cued}, that determined the extra input drive to neurons corresponding to the cued item and hence the ratio of cued/uncued stimulus-related activity in the normalized population. I obtained ML parameters of the model for each subject, allowing this weighting factor to take on a different value for each array size, while keeping the tuning curve width ω and population gain γ constant. Estimates of tuning width ($\omega = 0.49 \pm 0.05$) and population gain ($\gamma = 81 \pm 15$ Hz) were statistically indistinguishable from those obtained for the separate group of subjects in Experiment 1 ($t_{(13)} < 0.93$, $p > 0.37$). Curves in Figure 3*a* show the distribution of errors generated by the model with ML parameters, averaged across subjects. Lines in Figure 3*b* show error variance under the model.

Evidence for optimal weighting in human performance

Although previous studies found evidence for top-down control over memory resources favoring cued items, it has not been possible previously to assess whether this allocation was quantitatively matched to performance goals. In contrast, the constraints of the population coding model make it possible to identify optimal weightings of cued to uncued activity that would minimize expected errors on the task, given the frequency with which each class of item was tested.

To identify which weights were optimal for the task parameters of Experiment 2, I calculated the expected variance of errors on this task for a range of different values of the weighting factor α_{cued}. This also required specifying values of the population parameters ω and γ, and for this purpose I used the ML values obtained from observers in Experiment 1 (Fig. 2*b*). The values of

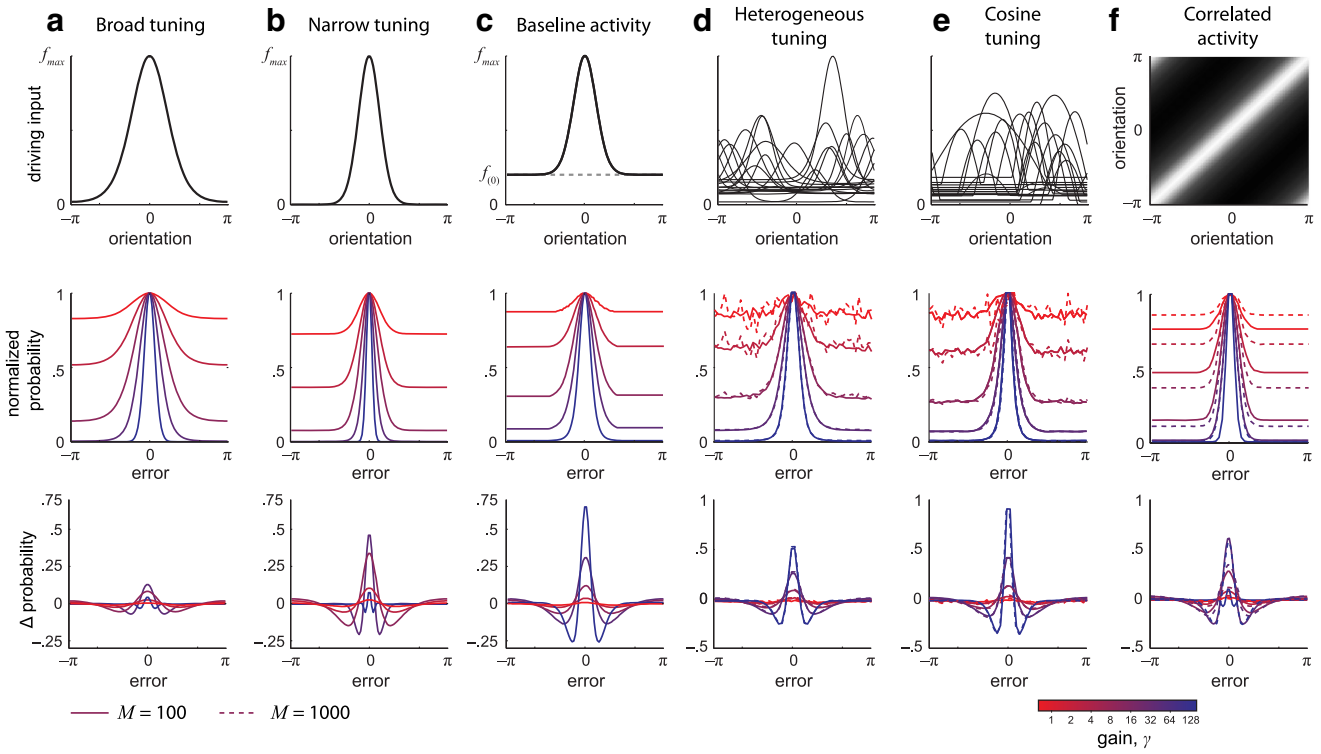

Figure 4. Effects of variations in tuning and noise correlations on decoding errors at low gain. ***a***, Simulated error distributions (middle, normalized by peak probability) and deviations of errors from the circular normal distribution (bottom) based on decoding a homogeneous population of neurons with broad tuning (top; $\omega = 0.5$). ***b***, Simulation results as in ***a*** but with narrower tuning curves (top; $\omega = 0.2$). Note that decoding precision is increased compared with a more broadly tuned population with the same gain (compare curves of same color in ***a*** and ***b***). ***c***, As in ***b*** but with the addition of a baseline level of activity (illustrated top; $f_{(0)} = 0.25$). For a specified gain, the presence of baseline activity decreases decoding precision. However, the pattern of deviations from normality is unaffected (bottom; see also Fig. 5). ***d***, Decoding errors obtained by simulation of neurons with randomly distributed preferred orientations and heterogeneous tuning curves (examples shown at the top; amplitude, $a = 1 \pm 0.5$; width, $\omega = 0.2 \pm 0.1$; baseline, $f_{(0)} = 0.25 \pm 0.125$). Heterogeneity had minimal impact on error distributions or deviations from normality, except to add noise at very low gains. ***e***, As in ***d*** but with cosine instead of bell-shaped tuning curves. ***f***, Decoding errors obtained from simulated populations with short-range pairwise correlations in spiking activity (illustrated top; $c_0 = 0.25$, other parameters as in ***b***). For a specified gain, decoding precision was decreased compared with uncorrelated activity. Unlike the other manipulations shown here (***a–e***), the impact of short-range correlations depended strongly on population size (solid lines, $M = 100$ neurons; dashed lines, $M = 1000$ neurons). However, the introduction of correlations had minimal impact on the pattern of deviations from normality (bottom).

α_{cued} that minimized error variance at each set size are shown as blue bars in Figure 3*d*.

These optimal weights were found to be greater than unity for all set sizes, indicating that an ideal observer should increase input drive to the cued item to store it with greater precision than uncued items. Furthermore, although relative test probability for cued and uncued items was held constant across set sizes (in the ratio 3:1), in all cases the optimal weighting factor α_{cued} increased as the number of uncued items grew (two vs four items: $t_{(7)} = 3.7$, $p < 0.01$; four vs eight items: $t_{(7)} = 3.1$, $p = 0.017$). So, minimizing error variability at higher memory loads required an increasing bias of coding activity toward the cued item, although its relative probability of being chosen for recall was unchanged.

White bars in Figure 3*d* show ML values of the weighting factors obtained from subject data in Experiment 2. In agreement with optimal predictions, these empirical α_{cued} values increased significantly with each increase in set size (two vs four items: $t_{(6)} = 2.8$, $p = 0.03$; four vs eight items: $t_{(6)} = 3.4$, $p = 0.01$) and were statistically indistinguishable from optimal values at every set size (two items: $t_{(13)} = 0.58$, $p = 0.56$; four items: $t_{(13)} = 0.23$, $p = 0.82$; eight items: $t_{(13)} = 1.8$, $p = 0.094$).

Because the optimal weights were calculated without fitting data from the cued task (i.e., by using ω and γ values from Experiment 1), the close correspondence with empirical weights obtained from Experiment 2 provides strong evidence that optimality principles drive performance on the task. However, I would also predict that each observer's weighting factors should be tuned to their own individual population parameters, and this was confirmed by comparing the empirical estimates of α_{cued} with optimal values calculated on the basis of ML estimates of ω and γ for the same observer (no significant difference at any set size: $t_{(6)} < 1.2$, $p > 0.15$; optimal and empirical weights obtained for each observer and set size were significantly correlated: $r = 0.46$, $p = 0.037$).

Comparing the recall precision (inverse variance) predicted under optimal weighting with that of empirically estimated weights showed that observers attained on average 95% of their theoretical maximum precision (two items: 95 ± 3%; four items: 96 ± 2%; eight items: 93 ± 3%).

Modifications and extensions to the model

The population coding model described above provides a remarkably parsimonious account of human recall performance based only on established principles of neural coding of sensory parameters. However, to limit the number of free parameters and for reasons of computational tractability, the model includes a number of simplifying assumptions that may limit its generality. Therefore, it is important to demonstrate that the model, and in particular the key observation that errors deviate from normality as gain decreases (Fig. 4*a*,*b*), is robust against modifications of these assumptions. Also, the estimates of population gain obtained by fitting the model to data are surprisingly low (on the

order of 10^2 spikes/s total for the population), and it is important to establish whether these low firing rates are critical to the success of the model or merely a consequence of choices made in its definition. For example, it should be noted that the length of the decoding window, which for illustration I set to 100 ms, is in a reciprocal relationship with the population gain: halving the decoding window to 50 ms would simply have doubled the estimates of gain while leaving the fit of the model to data unchanged.

One important consideration concerns interactions between the multiple orientations presented in each memory array. For simplicity of modeling, I assumed that each orientation was encoded by an independent subpopulation and that recall was based on decoding a unique subpopulation identified by the probe location. However, previous behavioral studies involving similar tasks have shown that responses are not entirely independent of unprobed stimuli (Bays et al., 2009; Gorgoraptis et al., 2011). In addition to variability in recall of the stimulus, errors may sometimes occur in identifying which of the stimuli in memory corresponds to the probe, with the result that an observer incorrectly reports a stimulus value corresponding to one of the other, unprobed items in the memory array.

I modified the basic population coding model to incorporate a probability of reporting a stimulus value decoded from a subpopulation corresponding to one of the unprobed locations. As expected, taking into account the influence of unprobed items in this way improved the fit of the model to data from Experiment 1 (log Bayes factor, 4.0 ± 2.4; BIC difference, 3.6 ± 4.4; AIC difference, 8.4 ± 4.4) with ML parameters corresponding to an ~3% probability of erroneous report per nontarget item ($\epsilon = 0.03 \pm 0.01$; $\omega = 0.67 \pm 0.08$; $\gamma = 165 \pm 37$ Hz). This is consistent with estimates from previous studies and confirms that unprobed response errors make a small but significant contribution to overall error variability. Future extensions of the population coding model could seek to incorporate these errors more naturally by relaxing the independence of subpopulations encoding different stimuli; for example, by adding a continuous spatial component to their tuning functions.

The present results were based on simulating a homogeneous population of neurons with uncorrelated activity, identical bell-shaped tuning functions, evenly distributed preferred stimulus values, and no spontaneous activity. In reality, sensory neurons typically exhibit short-range noise correlations and spontaneous baseline activity levels and vary randomly in their preferred stimulus values, as well as in their tuning curve shapes, widths, and amplitudes (Zohary et al., 1994; Ecker et al., 2010). Figure 4*c–f* plots results of simulations in which each of these simplifying assumptions was modified. Introducing a baseline level of activity (Fig. 4*c*) increased the variability in responses for a given population gain and tuning width (compare with curves of matching color in Fig. 4*b*). Introducing random variability into neural tuning functions had the effect of adding random structure to the response distributions at very low gains but minimal impact on the distribution of errors (Fig. 4, compare *d* with *c*). Modifying the shape of the tuning functions [from circular normal (Fig. 4*d*) to cosine (Fig. 4*e*)] also had negligible effect on decoding errors. Introducing short-range correlations between neurons (Fig. 4, compare *f* with *b*) increased variability for a given gain and tuning width.

In the homogeneous model, error distributions were independent of the size of the neural populations encoding each item. This was attributable to a continuous approximation to the distribution of preferred stimuli that simulations showed to be appropriate even for very small populations (e.g., >50 neurons) and the fact that the total activity in the population was fixed by the population gain γ. This was also true for the model with baseline activity (assuming this activity level was fixed as a fraction of the peak response of a neuron). Introducing heterogeneity in tuning curves could theoretically have introduced an effect of population size, but comparing simulations with 100 (Fig. 4*d*,*e*, solid lines) and 1000 (dashed lines) neurons found minimal effect on error distribution. However, the effect of short-range correlations between neurons was found to be strongly dependent on the size of the simulated population: introducing a fixed pairwise correlation between neurons with similar tuning produced a larger increase in variability for a population of 1000 (Fig. 4*f*, dashed line) than 100 (solid line) neurons (for full details, see Materials and Methods).

Critically, none of these modifications changed the fundamental pattern observed for the uncorrelated homogeneous population: in every case, decreasing population gain resulted in both an increase in variability and significant deviations from normality in the distribution of errors (Fig. 4, bottom row).

The addition of a baseline level of activation to the idealized population had a strong and consistent effect on response variability (Fig. 4*c*). To investigate this effect in more detail, I examined the ability of this modified model to fit experimental data from Experiment 1, initially treating baseline activity level as a free parameter.

The best fit to the data was obtained with a baseline corresponding to 7.6 ± 3.2% of peak activation. However, formal model comparisons favored the model without baseline activity (log Bayes factor, 3.7 ± 1.7), and closer examination indicated that the modified model was susceptible to overfitting. As shown in Figure 5*a*, the principal effect of increasing baseline activation was to shift ML estimates of population activity to higher values, without meaningfully altering the quality of the fit to empirical data [Fig. 5c; there was also a smaller effect on tuning width (Fig. 5*b*)]. So the present data are consistent with a model in which decoding is based on a small number of spikes, all of which code for the stimulus, but it is equally consistent with decoding based on a much larger number of spikes in which background firing makes up a substantial proportion of that activity.

This suggests that it is not the level of activity per se that determines decoding variability but rather the relative magnitude of signal-to-noise in that activity. Consistent with this interpretation, the SNR per neuron calculated from ML parameters was found to be approximately constant across changes in baseline activation (Fig. 4*d*; baseline 0% vs baseline 90% of peak activity: $t_{(7)} = 1.3, p = 0.23$). In other words, increases in noise attributable to stimulus-independent baseline activity could be countered by increases in gain, which boosted the strength of the coding signal. When the changes in signal and noise were matched, error variability and distribution were unchanged (Fig. 4*c*). I also found the SNR to have a simple relationship with memory load in the population coding model, being inversely proportional to the number of items in memory (see Materials and Methods).

It may prove possible to account for the effects of interneuronal correlations on model predictions (Fig. 4*f*) in a similar way: by considered their effects on the SNR calculated at the level of the population rather than individual neurons (Zohary et al., 1994). However, the effects of correlations on a population code are known to be dependent on details of the correlation structure (Averbeck et al., 2006) and to interact with other factors, including tuning heterogeneity and (as observed here) population size (Ecker et al., 2011). The computational demands of simulating large correlated populations make a full exploration of the con-

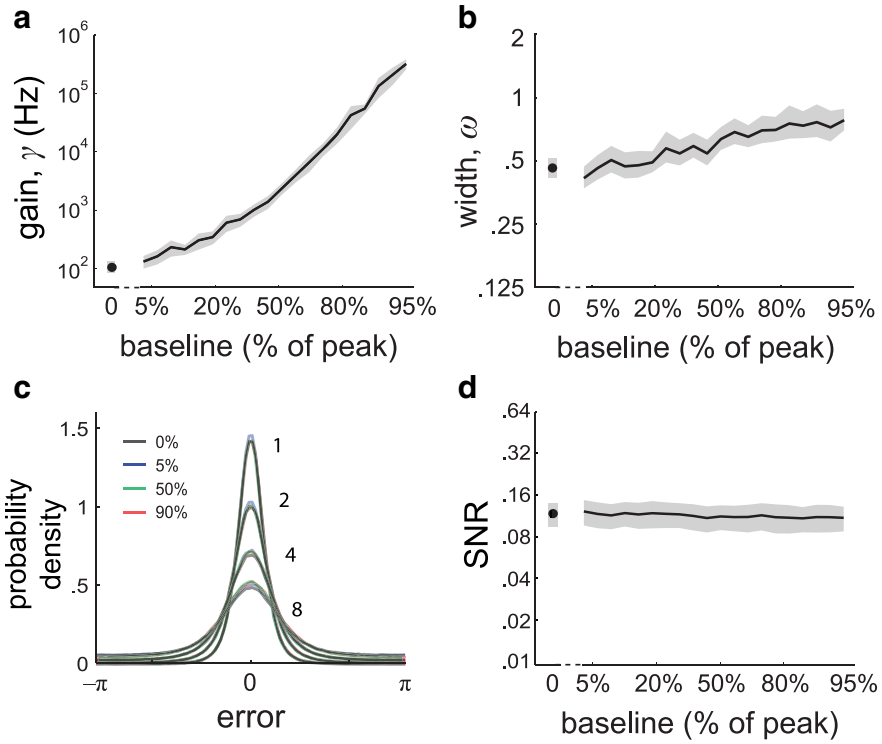

Figure 5. Effects of baseline activity on ML parameters and decoding errors. ***a***, Population gain (ML values for subjects in Experiment 1) as a function of baseline activation of the neural population (shaded area indicates ±1 SE). Note that the population gain that best fits the data rises steeply as the proportion of activity attributable to baseline increases. ***b***, ML values of the tuning width ω as a function of baseline activation. ***c***, Error distributions corresponding to ML parameters of the model with different baseline activation levels (1–8 items). It is not possible to reliably distinguish on the basis of error distributions between decoding of a population with low activity of which little or none is attributable to baseline (e.g., black, blue curves) and a population with substantially higher activity of which a much larger proportion is attributable to baseline (e.g., green, red curves). ***d***, SNR per neuron as a function of baseline activation, calculated from ML parameters. SNR shown is for a population encoding a single stimulus. The SNR corresponding to observed error distributions is approximately independent of the baseline activity level.

sequences for error distributions, or fitting such a model to behavioral data, impractical using the present methods.

Comparison with discrete and variable resource-based models

I compared the population coding model to two previously described models of memory errors, both based on distribution of a limited representational medium or resource. In the slots + averaging model (Zhang and Luck, 2008), the representational medium is subdivided into a small number of indivisible quanta, or slots, distributed between items in whole units. The model assumes that error in each memory representation has a circular normal distribution, and recall estimates are obtained by averaging over representations.

Figure 6*a–d* shows predictions of the slots + averaging model based on ML parameters obtained from the data in Experiment 1 ($M = 4.0 \pm 0.63$; $\sigma_1 = 0.63 \pm 0.05$). As the number of items increases, the mean number of memory quanta allocated to each item declines, providing a good quantitative fit to the observed changes in recall variance (Fig. 6*c*). However, the model does not provide a good account of the deviations from normality observed in empirical data, predicting large deviations (positive kurtosis) only for supra-capacity arrays (Fig. 6*d*).

In particular, I observed substantial deviations from normality in recall for a single visual feature (Fig. 2, left column), in which the slots + averaging model predicts errors will be normally distributed, with kurtosis around 0. A formal model comparison confirmed that the population coding model provided a significantly better fit to the experimental data than the slots + averaging model (log Bayes factor, 11.8 ± 4.3; BIC/AIC difference, 22.8 ± 8.6; model comparison favored the population coding model in seven of eight subjects).

The slots + averaging model can be viewed as a two-stage hierarchical model. The first stage determines the number of slots allocated to the probed item and hence the precision with which the item is stored. The second stage consists of drawing a random sample from a normal distribution with the specified precision: the outcome determines the error in recall. In most cases, it is not possible to equally distribute slots between memory items; therefore, the first stage (as well as the second stage) is probabilistic: the probed item will be stored with different levels of precision from trial to trial.

Figure 6*a* plots the cumulative distribution over precision predicted by the slots + averaging model with ML parameter values obtained for a representative subject. Because slots are allocated in discrete quanta, this cumulative distribution consists of a series of steps: each step corresponding to a different number of slots allocated to the probed item (here I assumed that slots are allocated as evenly as possible between items, but other allocation policies are possible and would predict different distributions over precision).

The recall advantage for a cued item seen in Experiment 2 has a possible explanation within the slots + averaging model: more slots could be allocated to the cued item than to uncued items, increasing the precision of its storage. However, model comparison again showed that this account (ML parameters: $M = 5.3 \pm 1.0$; $\sigma_1 = 0.89 \pm 0.12$; $A_{\text{cued}}^{N=2} = 3.3 \pm 0.68$, $A_{\text{cued}}^{N=4} = 2.0 \pm 0.54$, $A_{\text{cued}}^{N=8} = 1.7 \pm 0.56$) provided a significantly poorer fit to the data than the population coding model (log Bayes factor, 26.3 ± 8.0; BIC/AIC difference, 45.6 ± 16.4; model comparison favored population coding model in seven of eight subjects).

An alternative account for deviations from normality (van den Berg et al., 2012) is based on distribution of a continuous, rather than quantized, memory resource. In this model, the amount of resource the probed item receives, and hence the precision with which it is stored, varies randomly from trial to trial around a mean determined by the number of items. The model specifies a gamma distribution over precision to describe this variability (Fig. 4*e*). Like the slots + averaging model, this "variable precision" model assumes that the recall error is determined by drawing a random sample from a normal distribution with the specified precision.

Figure 6*e–h* shows predictions of the variable precision model based on ML parameters obtained from the data in Experiment 1 ($\bar{J}_1 = 17.6 \pm 3.1$; $\alpha = 1.36 \pm 0.09$; $\tau = 5.0 \pm 0.6$). The model provided a significantly better account of the data than the slots + averaging model (log Bayes factor, 12.7 ± 4.2; BIC difference, 19.8 ± 8.2; AIC difference, 24.6 ± 8.2; model comparison favored the variable precision model in six of eight subjects),

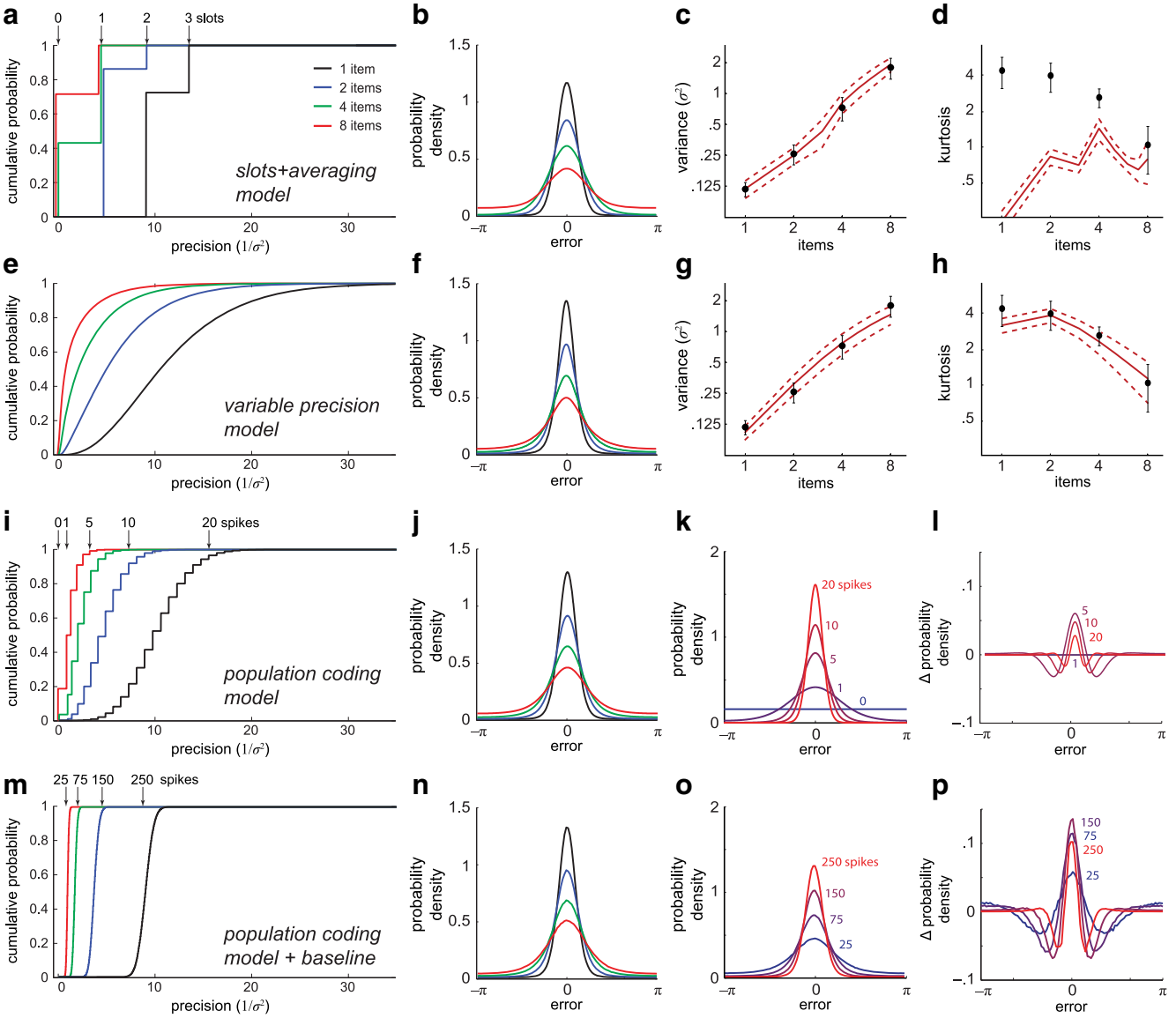

Figure 6. Comparison with resource-based models of working memory errors. ***a***, Errors under the slots + averaging model are drawn from a finite mixture of normal distributions of different widths, corresponding to different numbers of slots allocated to the probed item, and plotted here as a cumulative probability over precision (based on ML parameters for representative subject; Fig. 2*a*). Different set sizes are indicated by different colors (overlying lines are slightly shifted horizontally for ease of viewing). Note that, for large set sizes, there is a non-zero probability that an item receives zero slots, in which case errors are drawn from a distribution with zero precision, i.e., the uniform distribution. ***b***, Error distributions corresponding to the slots + averaging model parameters illustrated in ***a***. ***c***, Variance of error in recalled orientation for human observers (black symbols) and for the slots + averaging model with ML parameters (red curve, dashed lines indicate ±1 SE). ***d***, Kurtosis of recall errors for subjects (black symbols) and the slots + averaging model (red curve). ***e–h***, The variable precision model. Errors are drawn from an infinite mixture of normal distributions of different widths, shown here as a cumulative probability over precision (***e***). Corresponding error distributions are shown in ***f***; variance and kurtosis corresponding to ML parameters in ***g*** and ***h***. ***i–l***, In the population coding model, the precision of recall is correlated with the total number of spikes contributing to the estimate. The cumulative probability and corresponding precision of different spike counts is shown in ***i***, for the model with zero baseline activity. Error distributions are shown in ***j***. Note that, for large set sizes, there is a non-zero probability of zero spikes occurring within the decoding window, resulting in an estimate with zero precision, i.e., uniformly distributed. The error distributions (***k***) corresponding to a particular precision in ***i*** are approximately normally distributed: deviations from circular normal are plotted in ***l***. ***m–p***, Corresponding results for the population coding model incorporating baseline activity (example shown, 50% of peak). Spike counts corresponding to a specified precision (***m***) are substantially higher than in the zero baseline model, and there is negligible probability of zero spikes; errors corresponding to a specified spike count (***o***) deviate strongly from normal (***p***); the resulting error distributions (***n***) are indistinguishable from the zero baseline case (***j***).

including the changes of kurtosis with set size (Fig. 6*h*). Comparison with the population coding model did not find consistent differences between the models in their ability to fit data from Experiment 1 [log Bayes factor, −0.83 ± 1.6; BIC difference, 3.0 ± 3.4; AIC difference, −1.8 ± 3.4; model comparison favored population coding model (positive values) in four (AIC), five (Bayes factor), or six (BIC) of eight subjects].

The variable precision model does not describe how precision would be distributed under unequal allocations of resource and, in particular, the consequences for uncued items of increasing precision for a cued item are currently undefined, so it was not possible to obtain predictions under the model for Experiment 2. As for the population coding model, incorporating reports of unprobed items (see above) improved the fit of both slots + averaging and variable precision models to data, but the conclusions of model comparison were unchanged.

I examined whether errors generated by the population coding model, like the two previous models, could be described in

terms of a distribution over precision. Unlike the slots + averaging and variable precision models, which both incorporate two independent sources of noise in the generation of recall errors, in the population coding model all errors arise from a single source of noise: Poisson variability in spike generation. Nonetheless, it may be possible to find a mathematical description of the non-normal output of the population coding model in terms of a mixture of normal distributions of different precisions, as in the other two models. In particular, I noted that the magnitude of the decoding error was correlated with the total number of spikes generated during the decoding window by simulated neurons encoding the probed item: considering only trials on which a certain number of spikes were generated, error variability was larger for smaller spike counts (Fig. 6*k*).

Figure 6*i* illustrates how this observation could lead to a distribution over precision: each step in the cumulative probability function corresponds to a different total spike count and hence a different precision. Because the distribution of individual spike counts is independent Poisson, the distribution of total spike counts is also Poisson. The distribution over precision indeed bears resemblance to that dictated by the previous models, particularly the variable precision model (Fig. 6*e*). Furthermore, although the distribution of error corresponding to a given spike count is in most cases not normal, the deviations from normality are relatively small (Fig. 6*k*,*l*).

Based on Figure 6*i*, it would be tempting to conclude that the integer nature of spiking is critical to the ability of the population coding model to reproduce deviations from normality seen in the data. In particular, at low gains, I observed a non-zero probability of zero spikes occurring during the decoding period (e.g., with probability 0.2 for the representative subject tested with 8 items, red line). However, as indicated above, these very low gains may be a consequence of the simplified model of neural populations used in simulation, with both spontaneous baseline activity (Figs. 4*c*, 5) and interneuronal correlations (Fig. 4*f*), potentially contributing to increases in the gain at which a particular error distribution is observed. Figure 6*m–p* shows an example of results obtained from the version of the model incorporating baseline activity ($f_{(0)} = 0.5$; ML parameters of ω and γ for representative subject). As described above, this instantiation of the model produces fits to data that are indistinguishable from the zero-baseline model (Fig. 6, compare *n* with *j*) but at substantially higher gains. Here the probability of obtaining zero spikes is negligible, and, for a given population gain, there is relatively little variation in precision between trials with low and high spike counts (note steeper curves in Fig. 6*m* compared with *i*).

As shown in Figure 6, *o* and *p*, the distribution of errors corresponding to a given spike count is strongly non-normal for populations of neurons with significant baseline activity. An examination of spike counts in the simulations of heterogeneous and correlated populations (Fig. 4*d*,*f*) indicated that each of these modifications also had the effect of increasing non-normality of error distributions corresponding to a given spike count.

These results suggest a degree of mathematical similarity between the population coding model and previous models based on mixtures of normal distributions, which may provide an explanation at the neural level for the success of these cognitive-level models. However, the present evidence suggests that the basis for this correspondence in the total spike count may not extend straightforwardly to more realistic populations, composed of correlated neurons with spontaneous activity and varying tuning shapes and widths.

Additionally, both slots + averaging and variable precision models assume that the precision of a memory representation is a stable property determined at encoding, whereas the total spike count, like individual spike counts, is a random variable that varies from observation to observation. Recent results indicated that observers possess some knowledge of their uncertainty about a remembered stimulus on a trial-by-trial basis (Fougnie et al., 2012; Rademaker et al., 2012). Within the population coding model, this might correspond to knowledge of the total number of spikes on which a decoded stimulus estimate is based. However, it should be noted that an ideal observer would consider which neurons generated the spikes (and hence the full likelihood function), rather than just the total spike count, in assessing reliability.

Discussion

I have shown that a population coding model of the kind developed to explain responses to immediate sensory stimulation (Pouget et al., 2000) provides an accurate account of errors in short-term memory. In particular, deviations from the normal distribution that are characteristic of memory errors (Bays and Husain, 2008; Zhang and Luck, 2008; Fougnie et al., 2012; van den Berg et al., 2012) were found to be a predictable and general outcome of decoding stimulus parameters stored in the activity of tuned neurons. Critically, these deviations only became significant when the ratio of neural signal-to-noise was low and estimation variability consequently high.

The effect of increasing memory load on recall errors was successfully explained by a decrease in the gain of neural activity in inverse proportion to the number of items (Fig. 2). This is consistent with single-neuron and fMRI studies that have observed decreases with memory load in the information about a stimulus carried by neural activity (Buschman et al., 2011; Emrich et al., 2013). In the present model, the decrease in gain was implemented by divisive normalization, an established principle of neural processing whereby individual neuronal responses are divided by the summed activity of a neural population (Carandini and Heeger, 1994, 2012; Olsen et al., 2010; Ohshiro et al., 2011). The result was that the total excitation of neurons representing all memory stimuli was constant, providing a putative biological basis for a limited working memory "resource" (Palmer, 1990; Wilken and Ma, 2004; Bays and Husain, 2008; for a similar proposal in the context of multiple object tracking, see Ma and Huang, 2009; for a related hypothesis for temporal sequence memory, see Bradski et al., 1994).

In a normalized population, the effects of attention on sensory responses can be modeled as a multiplicative increase in the input drive to neurons encoding the attended stimulus or location (Reynolds and Heeger, 2009). Here I found evidence that a similar mechanism underlies observers' ability to control the contents of working memory. A cued item with higher probability of test was recalled with greater precision than other items, consistent with previous demonstrations of recall advantages for salient or behaviorally relevant stimuli (Bays and Husain, 2008; Shao et al., 2010; Bays et al., 2011a; Gorgoraptis et al., 2011; Melcher and Piazza, 2011). Differences in the variability and distribution of recall errors for cued and uncued items were successfully reproduced by weighting of the cued representation by a multiplicative gain factor (Fig. 3).

The population coding model allowed me to make quantitative predictions regarding the optimal weighting of stimuli in memory that would minimize errors on the task. For a fixed ratio of cued/uncued test probability, I found that observers should

increasingly weight memory representation in favor of the cued stimulus as load increased. This prediction was borne out in the distributions of observers' errors, demonstrating that human observers can fine-tune the contents of working memory to attain behavioral goals.

An important concept in models based on divisive normalization is that of the normalization pool, the pool of neurons whose summed activations provide the denominator of the divisive operation (Carandini and Heeger, 1994, 2012). In the model considered here, every memory neuron was a member of the normalization pool for every other neuron, with the result that every stimulus competed on an equal basis for representation in the population. Although this simple model proved very effective at accounting for human performance in the present experiments, it should be noted that divisive normalization operating across populations widely separated in space has not yet been identified experimentally. Note also that the model of divisive normalization used here is abstract: any realistic network implementation of normalization would be expected to itself suffer the consequences of spiking variability. One possible outcome is correlations in error magnitude across items on a single trial.

Future work could also consider the composition of the normalization pool in more detail. For instance, it is well established that competition for working memory storage is weaker between features from different feature dimensions (e.g., orientations vs colors) than within a dimension (Olson and Jiang, 2002; Wheeler and Treisman, 2002; Fougnie et al., 2010; Bays et al., 2011b); this could be a consequence of normalizing the activation of a neuron by a pool of neurons mainly selective for the same feature dimension. There is also some evidence that competition for working memory is weaker within than between visual hemifields (Umemoto et al., 2010; Buschman et al., 2011), which could reflect a greater contribution of within-hemisphere than between-hemisphere neuronal connections to the normalization computation.

In common with many previous theoretical studies of population coding (Snippe and Koenderink, 1992; Seung and Sompolinsky, 1993; Wilke and Eurich, 2002), the model of neural responses used here incorporated simplifying assumptions that are at best approximations to how real neuronal populations behave. This simple model had two key advantages that made it highly suitable for fitting to experimental data: (1) it was relatively tractable, both analytically and computationally; and (2) it had very few free parameters. Investigating a number of modifications to this model that would bring it closer in line with the behavior of cortical neurons recorded *in vivo* confirmed that the strong deviations from the normal distribution that make the model such a good fit to human recall errors are mostly insensitive to factors such as heterogeneity in tuning functions, correlations between neurons, and spontaneous activity (Fig. 4). However, these results also urge some caution in interpreting the population parameters obtained from fitting the unmodified model to data.

Specifically, the levels of population activity in the model that reproduce human levels of variability in recall would appear to indicate decoding of a very small number of spikes (for an illustration, see Fig. 6*i*). Although this is consistent with the neurophysiological observation that individual neurons and even individual spikes may carry a great deal of information about sensory stimuli (Britten et al., 1992), the results of simulations shown in Figures 4 and 5 suggest that these very low population gains could be an artifact of the simplifying assumptions of the model: both the spontaneous activity and correlations observed in real neural populations could contribute to increases in the population gain corresponding to empirically observed levels of variability.

The finding that working memory representations closely resemble sensory representations with a lower SNR finds a parallel in recent fMRI studies that have successfully decoded visual memories from signals in early visual cortex (Harrison and Tong, 2009; Ester et al., 2013). Although the initial presentation of visual stimuli evokes strong activity in these regions, the delay-period signal on which decoding is based is much weaker and mostly independent of memory load (Emrich et al., 2013), as expected for a normalized population. However, it remains to be established whether working memory is subserved by the same neurons that encode immediate sensory representations or by an analogous mechanism in a separate network.

An important prediction of the population coding model is that the distribution of recall errors depends on the width of neuronal tuning curves. The tuning widths estimated from the present experiments correspond quite well to those observed in recordings from orientation-selective neurons in primary visual cortex (Ecker et al., 2010). I suggest that the pattern of errors in recall from working memory provides a means of probing the neural representation of different features; this approach could profitably be extended to multidimensional feature spaces, such as color or location.

The present model represents a steady-state approximation to sustained activity, in which mean firing rates are constant over time. I have not attempted to model the initial burst of activity typically seen in sensory neurons after stimulus onset, which could reflect the initial non-normalized input from earlier sensory areas. Furthermore, working memory representations are typically short-lived, decaying on a timescale of seconds (Brown, 1958; Vogels and Orban, 1986). This may be a consequence of drift, over time, in the patterns of neural activity encoding stimulus parameters (Compte et al., 2000; Burak and Fiete, 2012). Persistent activity in recurrent networks diffuses because of neural noise, and the rate of this diffusion is related to the accuracy with which the memory state can be decoded by an ideal observer (Burak and Fiete, 2012). So the same factors that influenced decoding precision in the present study could also affect the rate at which memories degrade over longer time periods.

One testable prediction of this account is that rate of decay should increase with memory load. A corollary is that boosting the gain of neurons encoding one particular stimulus during maintenance would protect that stimulus from decay, at the cost of more rapid decay for other stimuli—this might provide an explanation for the "retro-cue effect," i.e., the recall advantage seen for an item cued long after its presentation (Griffin and Nobre, 2003; Matsukura et al., 2007; Pertzov et al., 2012).

References

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. CrossRef Medline

Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. Nat Rev Neurosci 7:358–366. CrossRef Medline

Bays PM, Husain M (2008) Dynamic shifts of limited working memory resources in human vision. Science 321:851–854. CrossRef Medline

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. CrossRef Medline

Bays PM, Gorgoraptis N, Wee N, Marshall L, Husain M (2011a) Temporal dynamics of encoding, storage, and reallocation of visual working memory. J Vis 11:pii:6. CrossRef Medline

Bays PM, Wu EY, Husain M (2011b) Storage and binding of object features

in visual working memory. Neuropsychologia 49:1622–1631. CrossRef Medline

Berens P, Ecker AS, Gerwinn S, Tolias AS, Bethge M (2011) Reassessing optimal neural population codes with neurometric functions. Proc Natl Acad Sci U S A 108:4423–4428. CrossRef Medline

Bradski G, Carpenter GA, Grossberg S (1994) STORE working memory networks for storage and recall of arbitrary temporal sequences. Biol Cybern 71:469–480. CrossRef

Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. J Neurosci 12:4745–4765. Medline

Brown J (1958) Some tests of the decay theory of immediate memory. Q J Exp Psychol 10:12–21. CrossRef

Burak Y, Fiete IR (2012) Fundamental limits on persistent activity in networks of noisy neurons. Proc Natl Acad Sci U S A 109:17645–17650. CrossRef Medline

Buschman TJ, Siegel M, Roy JE, Miller EK (2011) Neural substrates of cognitive capacity limitations. Proc Natl Acad Sci U S A 108:11252–11255. CrossRef Medline

Carandini M, Heeger DJ (1994) Summation and division by neurons in primate visual cortex. Science 264:1333–1336. CrossRef Medline

Carandini M, Heeger DJ (2012) Normalization as a canonical neural computation. Nat Rev Neurosci 13:51–62. CrossRef Medline

Compte A, Brunel N, Goldman-Rakic PS, Wang XJ (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cereb Cortex 10:910–923. CrossRef Medline

Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS (2010) Decorrelated neuronal firing in cortical microcircuits. Science 327:584–587. CrossRef Medline

Ecker AS, Berens P, Tolias AS, Bethge M (2011) The effect of noise correlations in populations of diversely tuned neurons. J Neurosci 31:14272–14283. CrossRef Medline

Emrich SM, Riggall AC, Larocque JJ, Postle BR (2013) Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. J Neurosci 33:6516–6523. CrossRef Medline

Ester EF, Anderson DE, Serences JT, Awh E (2013) A neural measure of precision in visual working memory. J Cogn Neurosci 25:754–761. CrossRef Medline

Fisher NI (1995) Statistical analysis of circular data. Camridge, UK: Cambridge UP.

Fougnie D, Asplund CL, Marois R (2010) What are the units of storage in visual working memory? J Vis 10:27. CrossRef Medline

Fougnie D, Suchow JW, Alvarez GA (2012) Variability in the quality of visual working memory. Nat Commun 3:1229. CrossRef Medline

Gorgoraptis N, Catalao RF, Bays PM, Husain M (2011) Dynamic updating of working memory resources for visual objects. J Neurosci 31:8502–8511. CrossRef Medline

Griffin IC, Nobre AC (2003) Orienting attention to locations in internal representations. J Cogn Neurosci 15:1176–1194. CrossRef Medline

Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. Nature 458:632–635. CrossRef Medline

Lara AH, Wallis JD (2012) Capacity and precision in an animal model of visual short-term memory. J Vis 12:pii:13. CrossRef Medline

Ma WJ, Huang W (2009) No capacity limit in attentional tracking: evidence for probabilistic inference under a resource constraint. J Vis 9:3.1–30. CrossRef Medline

Macke JH, Berens P, Ecker AS, Tolias AS, Bethge M (2009) Generating spike trains with specified correlation coefficients. Neural Comput 21:397–423. CrossRef Medline

Matsukura M, Luck SJ, Vecera SP (2007) Attention effects during visual short-term memory maintenance: protection or prioritization? Percept Psychophys 69:1422–1434. CrossRef Medline

Melcher D, Piazza M (2011) The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. PLoS One 6:e29296. CrossRef Medline

Ohshiro T, Angelaki DE, DeAngelis GC (2011) A normalization model of multisensory integration. Nat Neurosci 14:775–782. CrossRef Medline

Olsen SR, Bhandawat V, Wilson RI (2010) Divisive normalization in olfactory population codes. Neuron 66:287–299. CrossRef Medline

Olson IR, Jiang Y (2002) Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. Percept Psychophys 64:1055–1067. CrossRef Medline

Palmer J (1990) Attentional limits on the perception and memory of visual information. J Exp Psychol Hum Percept Perform 16:332–350. CrossRef Medline

Pertzov Y, Bays PM, Joseph S, Husain M (2013) Rapid forgetting prevented by retrospective attention cues. J Exp Psychol Hum Percept Perform 39:1224–1231. CrossRef Medline

Pouget A, Dayan P, Zemel R (2000) Information processing with population codes. Nat Rev Neurosci 1:125–132. CrossRef Medline

Rademaker RL, Tredway CH, Tong F (2012) Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. J Vis 12:21. CrossRef Medline

Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61:168–185. CrossRef Medline

Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. Proc Natl Acad Sci U S A 90:10749–10753. CrossRef Medline

Shao N, Li J, Shui R, Zheng X, Lu J, Shen M (2010) Saccades elicit obligatory allocation of visual working memory. Mem Cognit 38:629–640. CrossRef Medline

Snippe HP, Koenderink JJ (1992) Discrimination thresholds for channel-coded systems. Biol Cybern 66:543–551. CrossRef

Umemoto A, Drew T, Ester EF, Awh E (2010) A bilateral advantage for storage in visual working memory. Cognition 117:69–79. CrossRef Medline

van den Berg R, Shin H, Chou WC, George R, Ma WJ (2012) Variability in encoding precision accounts for visual short-term memory limitations. Proc Natl Acad Sci U S A 109:8780–8785. CrossRef Medline

Vogels R, Orban GA (1986) Decision processes in visual discrimination of line orientation. J Exp Psychol Hum Percept Perform 12:115–132. CrossRef Medline

Wheeler ME, Treisman AM (2002) Binding in short-term visual memory. J Exp Psychol Gen 131:48–64. CrossRef Medline

Wilke SD, Eurich CW (2002) Representational accuracy of stochastic neural populations. Neural Comput 14:155–189. CrossRef Medline

Wilken P, Ma WJ (2004) A detection theory account of change detection. J Vis 4:1120–1135. CrossRef Medline

Xie X (2002) Threshold behaviour of the maximum likelihood method in population decoding. Network 13:447–456. CrossRef Medline

Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. Nature 453:233–235. CrossRef Medline

Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 370:140–143. CrossRef Medline