Stochastic sampling provides a unifying account of working memory limits

Sebastian Schneegans, Robert Taylor & Paul M Bays*

University of Cambridge, Department of Psychology, Cambridge, CB2 3EB, UK

Attempts to characterize the limits of human working memory have differed on whether internal representations are discrete or continuous, with models of each type competing to best capture the errors observers make in delayed reproduction of elementary stimulus features. Here we show discretization only weakly discriminates between models; the critical distinction is instead between deterministic (fixed) and stochastic (randomly varying) limits, with only the latter compatible with observed human performance and the underlying biological system. Reconceptualizing existing models in terms of sampling reveals strong commonalities between seemingly opposing accounts: adding stochasticity to a discrete model brings it into closer correspondence with theories of neural coding, and puts its quality of fit on a par with continuous models, but also eliminates the stability and dependencies between items implied by a fixed set of "slots". A probabilistic limit on the number of items successfully retrieved is an emergent property of stochastic sampling, with no explicit mechanism required to enforce it. These findings resolve discrepancies between previous accounts and establish a unified computational framework for further investigating working memory.

Working memory refers to the nervous system's ability to form stable internal representations that can be actively manipulated in the pursuit of behavioral goals. Our understanding of the structure of working memory has been shaped by the competition between different formal models that aim to explain behavioral data by implementing different conceptual ideas. A

*Email for correspondence: pmb20@cam.ac.uk.

classical view of visual working memory (VWM) held that it was organized into a limited number of memory slots, each capable of holding a single object [1, 2]. This model was subsequently modified to allow multiple slots to hold the same object and be combined on retrieval to achieve higher precision [3]. This "slots+averaging" model incorporated aspects of an alternative view, which holds that VWM is better conceptualized as a continuous resource that can be flexibly distributed between different objects or visual elements [4, 5], accounting for set size effects in analogue report tasks [6] (Fig. 1A) and flexibility in prioritizing representations [7]. Variable precision models [8, 9] additionally proposed that double stochasticity is a key feature of VWM representations needed to explain human performance. An alternative approach [10] sought to explain VWM errors from neural principles as decoding variability in population representations [11], with the limited memory resource equated to the total neural activity dedicated to storage.

Here we show that these influential accounts of VWM can all be interpreted in the common framework of sampling [12]. This reconceptualization reveals unexpected similarities between the models, and identifies which of the differences are critical for reproducing experimentally observed response errors.

We first show how a population coding model [10] can, with some simplifying assumptions, be reinterpreted in terms of sampling (Fig. 1B–E). We consider a mathematically idealized population of neurons encoding a one-dimensional stimulus feature, where the amplitude of each cell's activity is determined by its individual tuning function. Neurons are assumed to share the same tuning function, merely shifted so the peak lies at each neuron's preferred feature value φ_i:

$$f_i(x) = f(x - \varphi_i). \tag{1}$$

Discrete spikes are generated from the cells' activity via independent Poisson processes. By associating each spike with a value in stimulus space equal to the preferred feature value of the neuron that generated it (the principle of population vector decoding [13]), we can interpret population activity as a set of noisy samples of the true stimulus value. If we pick at random any spike generated by the neural population in response to a stimulus value θ, we can determine the probability that it was produced by a neuron with preferred feature value φ. If we assume dense uniform coverage of the underlying feature space by neural tuning curves, this yields a continuous probability distribution $p(\varphi)$ over the space of preferred feature values (Fig. 1C).

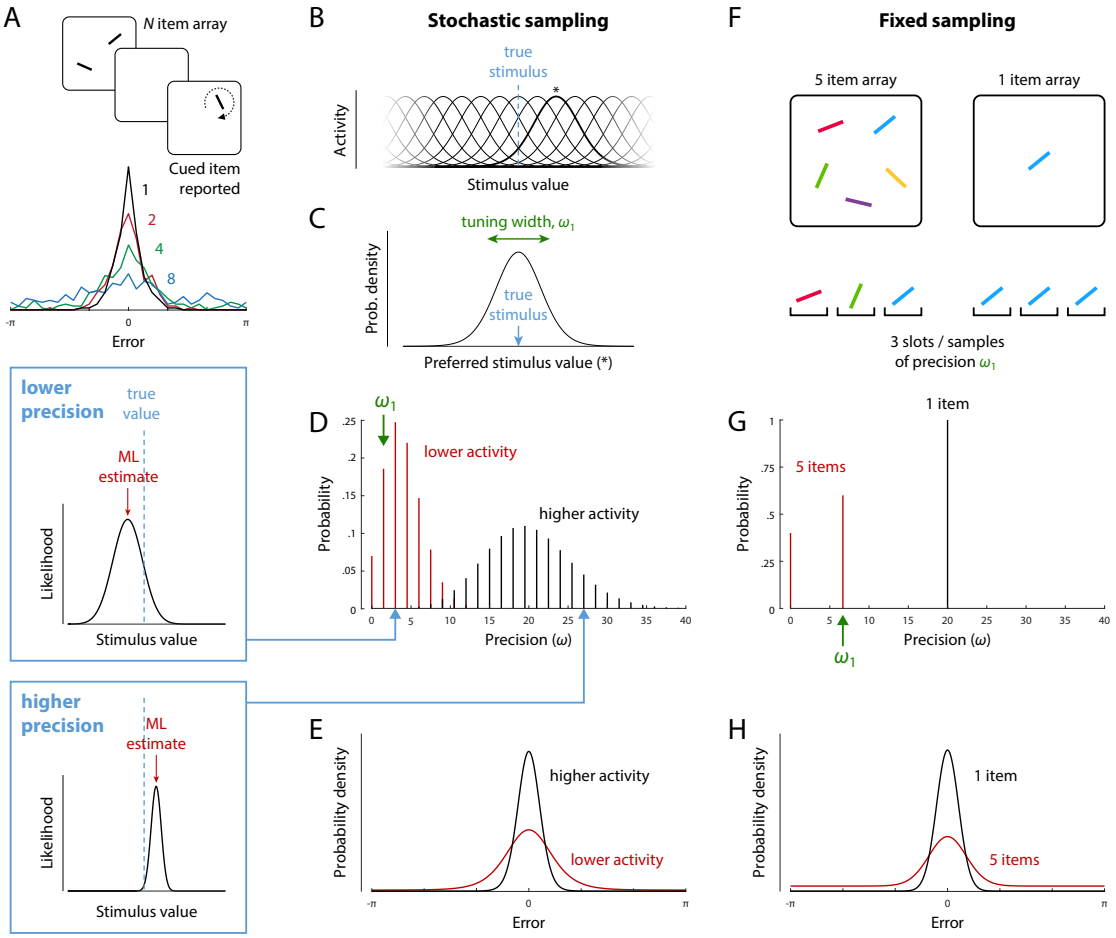

Figure 1: Working memory methods and models. (A) Analogue single-report task and distributions of response errors for a representative participant at different set sizes [10]. (B–E) A theoretical account of neural population coding can be re-interpreted as sampling. (B) The stimulus-invoked response of spiking neurons in an idealized population depends on their individual tuning (one neuron's tuning function and preferred value [*] is highlighted). (C) Probability distribution over stimulus space obtained by associating each spike with the preferred stimulus of the neuron that generated it. (D) Precision of maximum likelihood estimates based on spikes emitted in a fixed decoding window. Precision, defined as the width of the likelihood function (insets), is discretely distributed as a product of the tuning precision (ω_1) and the number of spikes, which varies stochastically. (E) Errors in estimation (here in a circular feature space) are described by a scale mixture of distributions with precision shown in (D). Assuming normalization of total activity encoding multiple items, larger set sizes correspond to less mean activity per item. (F–H) An account based on averaging limited memory slots can also be described as sampling. (F) Illustrates allocating a fixed number of samples or slots (here, three) to memory displays of different sizes. (G) Precision is discretely distributed as a product of the tuning width, ω_1, and the number of samples allocated per item. (H) Corresponding distribution of estimation errors.

This distribution has the same shape as the neural tuning curves and is centered on the true stimulus value:

$$p(\varphi) \propto f(\theta - \varphi) \tag{2}$$

Retrieval of a feature value is modeled as decoding of the spikes generated within a fixed time window. In the idealized case with Gaussian tuning functions, the maximum likelihood (ML) decoder generates an estimate by simple averaging of the spike values:

$$\hat{x}_{ML} = \frac{1}{n} \sum_{j}^{n} \varphi_{(j)} \tag{3}$$

where $\varphi_{(j)}$ is the preferred feature value of the neuron that generated the jth spike.

Due to the superposition property of Poisson processes, the number of spikes – or samples – generated by the neural population within the decoding window is also a Poisson random variable. If the total spike rate in the neural population is normalized, or fixed at a population level γ, it implements a form of limited resource [10]. If distributed equally between items, this results in a mean number of samples available for decoding each stimulus that is inverse to the set size N. This has been shown to quantitatively capture the set size effect in analogue report tasks (Fig. 1A). The actual number of samples available in this model for decoding each item in a single trial, n_k, is independently drawn from a Poisson distribution,

$$n_k \sim \text{Poiss}\left(\frac{\gamma}{N}\right). \tag{4}$$

The neural population model can therefore be interpreted as a *stochastic* sampling model.

The most prominent discrete model of visual WM, the slots+averaging model [3], can also readily be interpreted in terms of sampling (Fig. 1F–H). Each slot is postulated to hold a representation of a single object with a fixed precision, and thus provides a noisy sample of the objects' feature values (note that the slot concept is commonly associated with an object- rather than feature-based view of WM storage, but this distinction is not our present concern). Multiple slots, or samples, that correspond to the same object are averaged at retrieval to enhance the precision of the estimated stimulus feature. Thus, the format of representation and the decoding mechanism are identical to the stochastic sampling model. There is one critical difference, however: The slots+averaging model assumes that the total number of samples

available for representing multiple items is fixed, i.e.

$$\sum_{k}^{N} n_k = K \tag{5}$$

This has also been the most common assumption in previous sampling-based models in the attentional and memory literature [12, 14, 15] (but see also [16]). We will refer to this as a *fixed* sampling model.

Predictions for precision and error distributions

We now consider the distribution of representational precision in these models. For any given set of samples, the information they provide about the stimulus is described by the likelihood function, $\mathcal{L}(x|\boldsymbol{\varphi}) = p_x(\boldsymbol{\varphi}|x)$. The width of the likelihood function is a measure of uncertainty in the estimate: a set of samples with a broad likelihood function (Fig. 1, top inset) is compatible with many different feature values, whereas a narrow likelihood function (bottom inset) identifies a value more precisely. While a pattern of samples may have a sharp likelihood function with a peak far from the true estimate (a kind of "false alarm"), statistically this is unlikely.

If the sample values follow a normal distribution with variance σ^2 centered on the true stimulus value, then the likelihood function is also normal, with a width that depends only on the number of samples available for decoding,

$$\mathcal{L}(x|\boldsymbol{\varphi}) \propto \phi\left(x; \hat{x}_{ML}, \frac{\sigma^2}{n_k}\right) \tag{6}$$

Furthermore, for a specified number of samples, the ML estimate is distributed around the true stimulus value as a normal with the same width as the likelihood,

$$\hat{x}_{ML}|n_k \sim \mathcal{N}\left(x, \frac{\sigma^2}{n_k}\right). \tag{7}$$

This correspondence between the likelihood function and trial-to-trial variability is not universal, but does apply to all the models considered in this study (see SI), and justifies defining the precision (which we will denote ω) of an individual estimate in terms of the width of the likelihood function.

For the stochastic sampling model based on population coding, precision has a Poisson distribution (Fig. 1D), scaled by the precision of a single sample which is determined by the neural tuning function, $\omega_1 = 1/\sigma^2$,

$$\frac{\omega}{\omega_1} \sim \text{Poisson}\left(\frac{\gamma}{N}\right) \tag{8}$$

Example distributions of decoding error are shown in Fig. 1E, where we have made a transition from 1-D Euclidean to circular stimulus space, corresponding more closely to the feature dimensions (e.g. orientation, hue) commonly used experimentally. The distribution of errors can be described as a scale mixture of normal distributions with precision proportional to the sample count (due to the circular stimulus space, this is a close approximation rather than exact; see SI). The dispersion of errors increases with decreasing activity (e.g. as a result of increasing set size; black curve vs red curve in Fig. 1E) and the distribution deviates from normality, with this effect being particularly evident at lower activity levels (red curve) where long tails are observed.

For the fixed sampling model, making the common assumption that samples are distributed as evenly as possible among items [9, 17], we obtain a discrete distribution over at most two precision values (Fig. 1G), which are multiples of the precision of one sample, ω_1. As in the stochastic model, mean precision is inversely proportional to set size, but because the distributions over precision differ, the fixed and stochastic models make distinct, testable predictions for error distributions (Fig. 1H).

Behavioral response errors discriminate between models

We tested the ability of stochastic and fixed sampling models to capture response errors in analogue report tasks. We fit the models to a large dataset of single-report tasks originating from different labs (see SI) and also to a set of whole-report experiments [18], in which participants reported the feature values of all items in a sample array, either in a prescribed random order or in an order freely chosen by each participant on each trial. While only a single study, the whole-report results include information regarding correlations in errors between items represented simultaneously in WM that could differentiate the models. For all models, we assumed that participants gave their responses in order of decreasing precision (corresponding to decreasing number of samples) on free-choice trials.

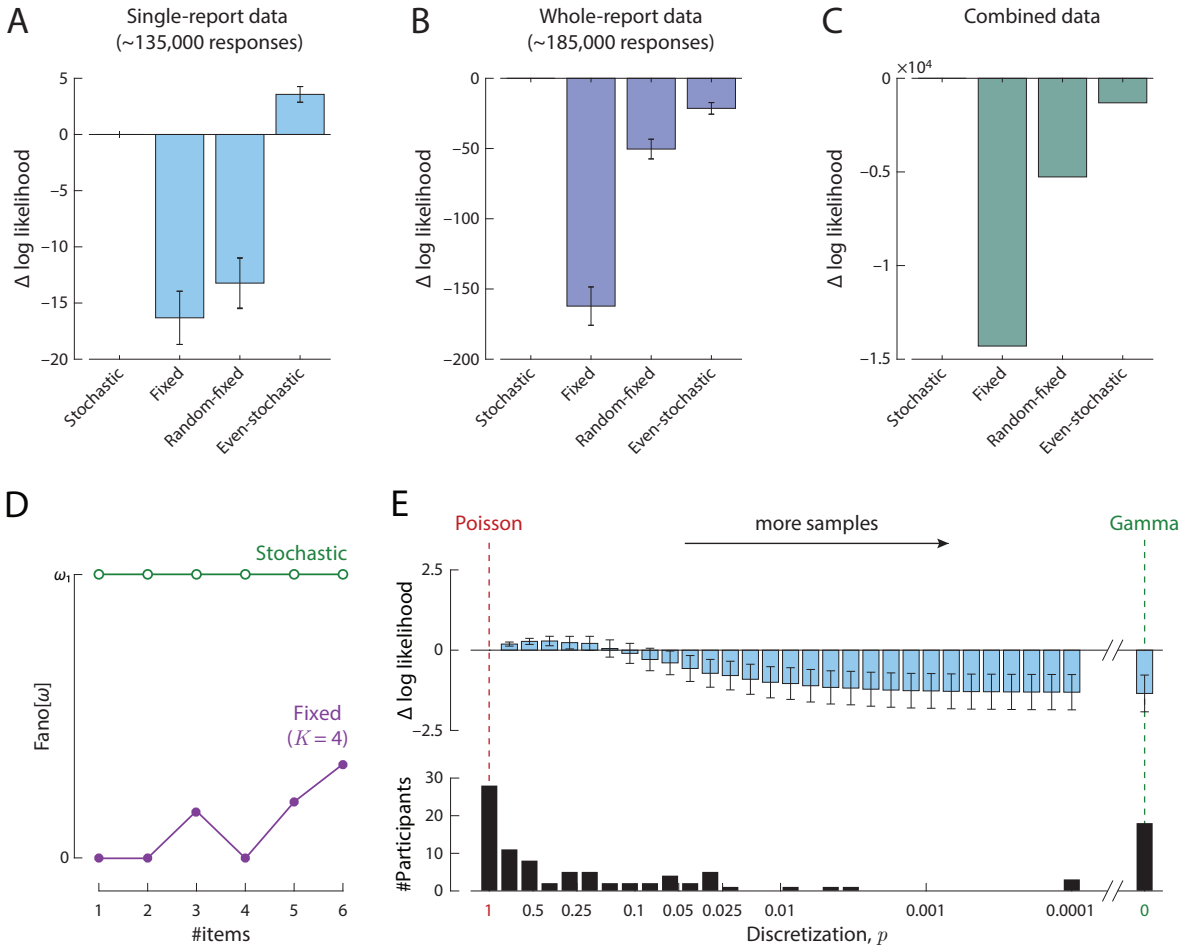

Figure 2: Model comparison based on single- and whole-report data. (A) Mean difference in log-likelihood of each model from the stochastic sampling model (with independence between items), for a benchmark data set of single-report experiments. More positive values indicate better fits to data. Errorbars indicate ± 1 SE across participants. (B) The same comparison for a set of whole-report experiments. (C) Total difference in log-likelihood between models across single- and whole report experiments. (D) Fano factor (ratio of variance to mean) of precision distribution. A constant Fano factor is characteristic of the stochastic model and contrasts with the varying Fano factor (dependent on set size and number of samples) in fixed sampling. (E) Mean difference in log-likelihood for differing levels of discretization in the generalized stochastic model. Differences are plotted relative to the maximum discretization ($p = 1$; left) corresponding to the standard stochastic model with Poisson-distributed precision. Lower discretization ($p < 1$) corresponds to more samples each of lower precision, converging to a continuous Gamma distribution over precision as p approaches zero (right). All models have the same number of free parameters and include a fixed per-item probability of swap errors (see SI).

Overall, the stochastic model fit data substantially better than the fixed sampling model for both single-report (Fig. 2A; difference in log likelihood per participant [ΔLL] = 16.3 $\pm$ 2.37 [M $\pm$ SE]) and whole-report tasks (Fig. 2B; ΔLL = 162 $\pm$ 13.6), indicating that stochasticity is critical for capturing behavioral performance. In contrast to previous interpretations [18], we did not find that the results of the whole report tasks support a slot-like mechanism with a fixed limit on the number of memorized items.

We tested two additional variants of the stochastic and fixed models set out above. In the *random–fixed* model, the total number of samples was fixed but distributed randomly between items. This model provided an improved fit to data compared to the fixed model with even allocation (moderately for single-report, ΔLL = 3.07 $\pm$ 1.10; strongly for whole-report, ΔLL = 112 $\pm$ 11.9), but was still substantially worse than the stochastic model in both cases (single-report, ΔLL = 13.2 $\pm$ 2.24; whole-report, ΔLL = 50.4 $\pm$ 7.03). In the *even–stochastic* model, the total number of samples was a Poisson random variable, but the samples were distributed as evenly as possible between items. This model achieved a better fit to single-report data than the stochastic model with independent sample counts for each item (ΔLL = 3.57 $\pm$ 0.697), but provided a substantially worse fit to whole-report data (ΔLL = 21.4 $\pm$ 4.12). Combining evidence across all participants and tasks, the stochastic model with independent sample counts was strongly preferred over this and the other alternative models (total ΔLL $>$ 1450; Fig. 2C).

Generalizing the stochastic model

One feature that strongly distinguishes the fixed and stochastic models is the relationship between mean and variability of the precision distribution. Because of its basis in independent Poisson events, the stochastic model has a constant Fano factor (the ratio of variance to mean), equal to the precision of a single sample, $Var[\omega]/E[\omega] = \omega_1$ (Fig. 2D, green). This is a direct consequence of the fact that the Poisson distribution itself has a Fano factor of one. In contrast, the Fano factor for the fixed sampling model is on average lower (mean $\sim$0.25 of ω_1 based on ML parameters and typical set sizes) and varies in an idiosyncratic manner between set sizes, due to the varying combinatorial possibilities of allocating a fixed number of samples to a fixed number of items (Fig. 2D, purple).

Fitted parameters of the stochastic sampling model indicate that estimates are based on a very

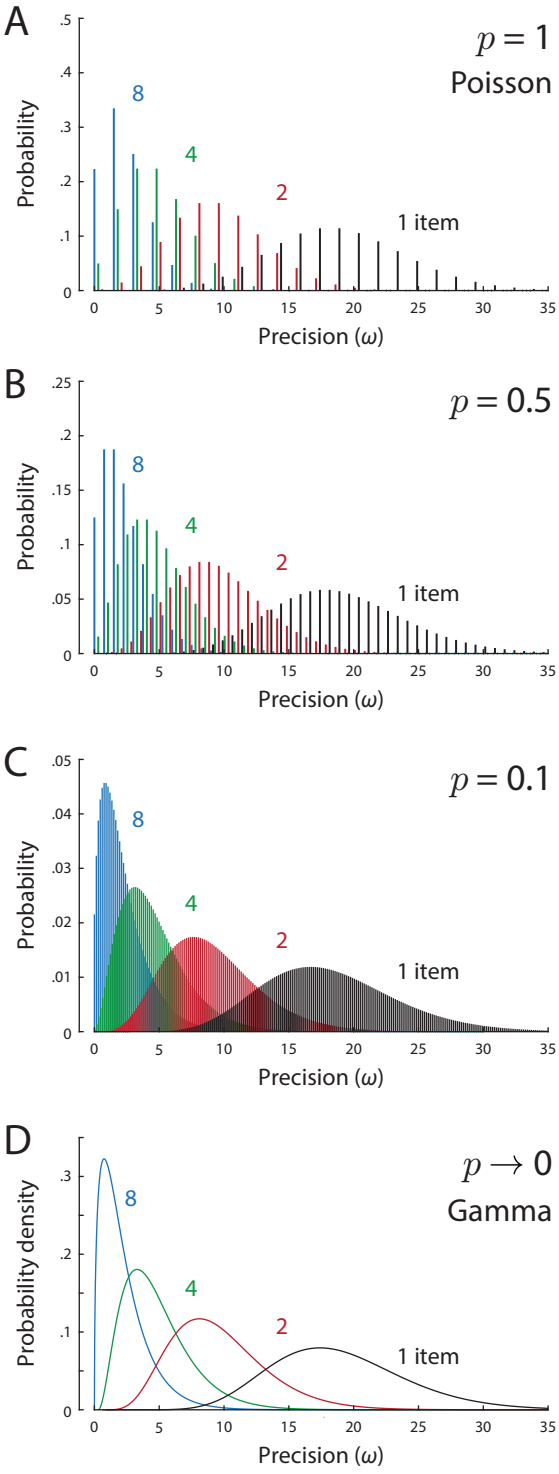

Figure 3: Precision distributions in the generalized stochastic model, for different levels of discretization, p, and different set sizes. (A) Example of discrete Poisson-distributed precision values (p = 1). For typical ML parameters, estimates are based on a small mean number of samples (here, γ = 12) each of moderate precision (ω_1 = 1.5). (B & C) With decreasing discretization ($p < 1$), estimates are based on larger mean numbers of samples and discrete precision values are more finely spaced. (D) In the limit as discretization falls to zero, the mean number of samples becomes infinite and the distribution over precision approaches a continuous Gamma distribution. The ratio of variance to mean precision (Fano factor) is fixed (at ω_1 = 1.5) across all set sizes and levels of discretization.

small average number of samples (e.g. mean of ∼13 samples based on fits to single-report data). From a neural perspective, this is a consequence of the idealized assumptions about population coding underlying the model, in particular that each spike from each neuron carries independent information about a stimulus (see Discussion below). The result is that precision of decoded estimates could take on only a limited set of possible values, and error distributions reflect a discrete mixture of distributions with different widths. To investigate whether discreteness and/or low numbers of samples are important for reproducing human performance, we implemented a generalization of the stochastic model in which the number of samples was free to vary while the Fano factor remained fixed. Specifically the distribution over precision was obtained as a scaling of the negative binomial distribution,

$$\frac{\omega}{\omega_1 p} \sim \mathrm{NegBin}\left(\frac{\gamma}{N}\frac{1}{1-p}, p\right). \tag{9}$$

The parameter p controls the discretization of the precision distribution: p = 1 corresponds to the Poisson model described above and illustrated in Fig. 3A (strictly Eq. 8 is the limit of Eq. 9 as $p \to 1$), while $p < 1$ corresponds to a stochastic model with a greater mean number of samples, $\bar{n} = \gamma/p$, each with a lower individual precision, $\omega_1 p$. The mean and variance in precision ($E[\omega] = \omega_1\gamma/N$, $Var[\omega] = \omega_1^2\gamma/N$) are independent of the discretization p, and the Fano factor is fixed at ω_1. Examples of precision distributions with different discretizations are shown in Fig. 3B & C.

As the discretization parameter becomes very small ($p \to 0$), the number of samples becomes very large and the distribution of precision described by Eq. 9 approaches a continuous function (Fig. 3D; see SI), specifically the Gamma distribution,

$$\omega \sim \mathrm{Gamma}\left(\frac{\gamma}{N}, \omega_1\right). \tag{10}$$

Two previous studies [8, 9] independently proposed that a continuous scale mixture of normal distributions with Gamma-distributed precision provided a good account of VWM data, but could not motivate this choice of distribution theoretically. In particular [9] proposed distributing precision as $\mathrm{Gamma}(\bar{J}_1/N^\alpha, \tau)$, with $\bar{J}_1$, τ and α free parameters. With α = 1 this is identical to Eq. 10 (see SI for results regarding this parameter). We can now explain Gamma-distributed precision as a limit case of the stochastic sampling model with large numbers of low-precision

samples.

Fig. 2E (top) shows the results of fitting the generalized stochastic model with different levels of discretization, p, to the single-report dataset. The best fit was obtained with a discretization roughly one third that of the Poisson model, p = 0.39. However, varying discretization produced differences in fit an order of magnitude smaller than those between fixed and stochastic sampling (varying by ∼1.5 versus ∼15 log likelihood points). Fitting the same model with p as a free parameter that could vary between participants, we found that ML estimates of discretization were very broadly distributed (Fig. 2E, bottom), with a majority of participants (72%) best described by a sampling model with less discreteness than the Poisson, and a minority (18%) better captured by the continuous limit ($p \to 0$) than any discrete value of p we tested (as low as 0.0001, corresponding to ∼100,000 samples). Formal model comparison was equivocal with respect to an advantage of including the discretization parameter in comparison to either the Poisson model (i.e. p = 1; ΔAIC = –0.61 ± 0.49; ΔBIC = +4.2 ± 0.46; negative values favor the added parameter) or the continuous Gamma model (i.e. $p \to 0$; ΔAIC = –3.3 ± 0.93; ΔBIC = +1.5 ± 0.89). Overall, these results do not allow strong conclusions to be drawn regarding the discreteness of sampling, which has relatively little effect on model predictions or the quality of fits.

Probabilistic item limits

In the fixed sampling model, at higher set sizes, a meaningful proportion of estimates are random "guesses" based on no samples (Fig. 4A & B). Specifically, if an estimate was generated for every item in the memory array, then as set size N increased, the number of estimates based on at least one sample, $S_{\omega>0}$, would reach a maximum at the fixed total number of samples,

$$\lim_{N\to\infty} S_{\omega>0} = K \tag{11}$$

This is a trivial consequence of sharing out a fixed number of samples evenly between items.

In the stochastic model with Poisson variability (p = 1), the number of samples available for each item varies probabilistically and independently of the other items. There is again a probability

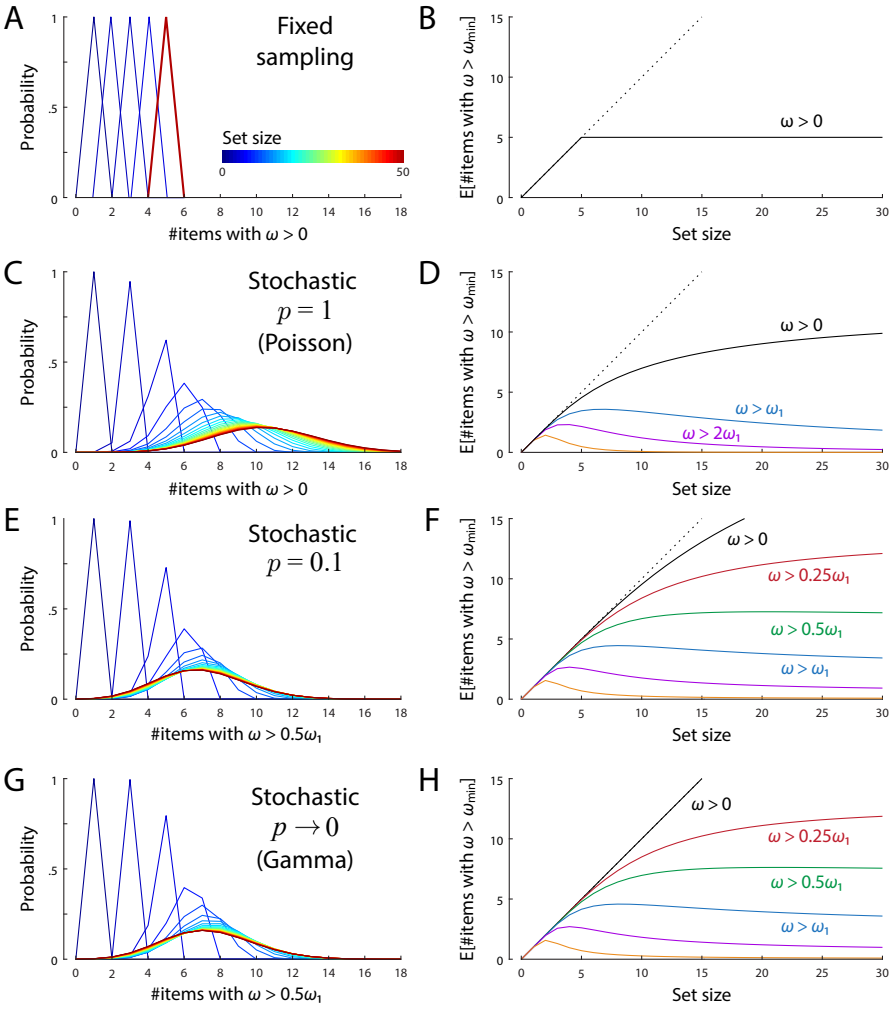

Figure 4: Item limits in sampling models. The left panel for each model shows how the probability distribution of the number of items recovered with greater than zero precision (A & C; greater than a fixed threshold for E & G) changes with set size (color coded, increasing blue to red; discrete probability distributions are depicted as line plots for better visualization). The right panels plot the mean number of items with above-threshold precision as a function of set size for different threshold values. Thresholds are defined as a proportion of the base precision ω_1. (A, B) In the fixed sampling model the number of items with non-zero precision increases with set size, then plateaus when the number of items equals the number of samples. (C, D) The stochastic sampling model with Poisson variability also has a limit on the number of items with non-zero precision, although this limit is probabilistic and emerges asymptotically (converging to the distribution shown by the red curve in C for large set sizes, corresponding to the mean number of items plotted as black curve in D). (E–H) Stochastic models with lower discretization (E & F) or continuous precision (G & H), display similar probabilistic item limits for precision exceeding a fixed threshold, but with the expected number of items saturating at different values depending on threshold (colors in right-hand plots).

of making an estimate based on zero samples,

$$\Pr(\omega = 0) = \exp\left(-\frac{\gamma}{N}\right), \tag{12}$$

and the number of non-random retrievals in a set of N items has a Binomial distribution,

$$S_{\omega>0} \sim \mathrm{Bin}\left(N, 1 - \exp\left(-\frac{\gamma}{N}\right)\right). \tag{13}$$

As set size increases, the mean number of estimates based on at least one sample reaches a maximum at the expected total number of samples (Fig. 4C & D). However, unlike the fixed sampling model, this limit is probabilistic and (as illustrated in Fig. 4C) the actual number will vary from one set of memory items to the next, converging to a Poisson distribution for large N,

$$\lim_{N\to\infty} S_{\omega>0} \sim \mathrm{Poisson}(\gamma). \tag{14}$$

As we increase the expected number of samples by reducing the discretization, p, the probability of zero samples falls to zero: $\Pr(\omega = 0) = p^{\gamma/(N(1-p))}$. However, qualitatively similar patterns are obtained by setting a threshold precision ω_{min} and counting the number of items that exceed this precision (illustrated for $p = 0.1$ in Fig. 4E & F). For small thresholds (on the order of the base precision ω_1), the expected number of items exceeding this precision saturates with increasing set size, although the specific number at which this function saturates depends on the threshold (different colors in Fig. 4F correspond to different threshold levels of precision). This pattern of probabilistic limits is largely unchanged by further decreases of discretization, up to and including the continuous limit $p \to 0$ corresponding to the Gamma model (Fig. 4G & H).

Discussion

Taking as a starting point a mathematical idealization of the way neural populations encode information, we have shown that retrieval of a visual feature from working memory can be described as estimation based on a stochastically varying number of noisy samples. By reconceptualizing alternative models within this same sampling framework, we were able to identify this stochasticity, rather than discreteness, as the critical factor determining their differing abil-

ities to capture behavioral data.

The slots+averaging model, because it modified the original slot model to allow multiple representations with independent noise, is equivalent to a sampling model with a fixed number of samples [14]. We have shown that adding stochasticity to the number of samples and randomizing their allocation to items makes this model's predictions effectively indistinguishable from an idealized population coding account, in the process removing the discrepancy in the models' ability to fit behavioral data. It should be noted that these modifications also sever the last remaining connections to the classical account of WM as consisting of a fixed set of memory slots. Indeed, in the best-fitting models, there is not only variability in the number of samples from trial-to-trial and item-to-item, but that variability is independent from one item to another: this is antithetical to the slot concept, whereby if a slot is occupied by one item, fewer slots are available for other items. In contrast to slots, the stochastic sampling account allows working memory to be described within the same computational framework as attention, and has a plausible link to neurophysiology in the form of population coding.

We have shown that variable precision models can also be described within the sampling framework, as the continuous limit approached as samples are made less precise and more numerous, while maintaining the fixed proportionality between the variance and mean of precision in the decoded estimate. The fully continuous model with Gamma-distributed precision proposed in previous studies provided fits to data that were overall a little worse than the discrete Poisson model, but when we attempted to fit discretization as a free parameter we found ML estimates varied widely between participants, with many best fit by continuous or near-continuous versions of the generalized model. We conclude that response error data does not provide evidence for any particular level of discreteness in representation, or any particular mean number of samples.

Here we have focused on experiments in which each item in an array is equally likely to be tested and so has equal priority for representation. However, the view of WM as a limited resource has also been motivated by studies showing that prioritized items can be represented more precisely in WM, at the cost of decreased precision for other items [7, 19–22]. The population coding model has previously been shown to reproduce data from tasks that encourage unequal allocation of resources (e.g. where a cue indicates that one item in a set has greater

chance of being tested). To understand how this result translates to a sampling framework it is necessary to make a distinction between two concepts that have previously been elided: discreteness of representations and discreteness of resources. In the stochastic sampling account the number of samples (discrete representations) that will be obtained for each item at retrieval is unpredictable, so giving an item increased priority at encoding cannot correspond to allocating more samples to it. Instead, the resource that is distributed between items according to their priority corresponds to the mean or *expected* number of samples, which is constant and continuous (there is no requirement for it to be an integer nor to be distributed in integer units). This distinction already exists in the population coding account, where the constant resource (γ in Eq. 4) is the sum of neurons' continuous-valued firing rates or – more concretely – membrane potentials, whereas the decoded estimate is based on the expression of this rate (or potential) in discrete spikes.

The idealized description of population coding on which we based the stochastic sampling model overlooks a number of important considerations in order to reveal relationships between cognitive and neural-level accounts of WM. In real neural populations there is considerable variability between cells in tuning width and maximum and minimum firing rates. Due to this heterogeneity, neurons differ in the amount of information each spike provides about a stimulus. From a sampling perspective, this means that estimates are based on samples that vary in precision, and this has the effect of "smoothing out" the discrete distribution of precision values predicted by the stochastic model. We found that adding heterogeneity of the kind observed in neural recordings further improved fits of the stochastic model to data (see SI).

Electrophysiological recordings show that individual neurons and even individual spikes are remarkably informative about visual stimuli [23, 24], rivalling the whole animal in accuracy of stimulus discrimination. This would seem to imply that human-level behavioral performance could be achieved on the basis of very low spike counts, and indeed this would be consistent with the small numbers of samples obtained in fits of the idealized model with Poisson variability. In reality, the spikes generated by a neural population are not independent events, but rather correlated within and between neurons. This will tend to result in deviations from the simple additivity assumed by sampling, although the exact consequences depend on details of the correlation structure that are difficult to measure experimentally [25–27], and suboptimal inference (in the form of a mismatched decoder) is likely to also play a part [28]. Decoding

properties of populations with these features are in general not amenable to mathematical analysis, but simulation studies (as in e.g. [10]) indicate that, except for occurring at higher spike counts, errors obtained from idealized populations may in many cases provide a good guide.

In keeping with most previous work on visual WM limits, we have not here attempted to reproduce the variations in bias and precision that are observed for different feature values, exemplified by the finding that cardinal orientations can be reproduced with greater precision than obliques. However, previous work has shown that these effects can be simply and elegantly captured within the population coding framework via the principle of efficient coding [29–31]. The idea is that neural tuning functions are adapted to the stimulus statistics of the natural environment in such a way as to maximally convey information in that environment (effectively by distributing neural resources preferentially to the most frequently-occurring stimulus features). Although it should be possible to formulate this model as a modification of stochastic sampling, without reference to neural populations, it seems that the modifications required would not have a natural explanation within the sampling framework. These observations, and the results of incorporating heterogeneity described above, illustrate the value of connecting abstract cognitive models to neural theory.

A recent proposal that WM errors can be explained in terms of a perceptual rescaling of stimulus space can also be expressed in terms of population coding, with some minor differences from the version presented here (see [32], for details and discussion). In particular, the idea of retrieval based on normally distributed “memory-match” signals maps exactly onto an idealized population code with continuous-valued activity and constant Gaussian noise [33, 34]. This predicts a continuous distribution over precision, not dissimilar to the Gamma distribution. Continuousness in representation does not appear a necessary component of this account, however, and it should be possible to reformulate it with arbitrary levels of discreteness, as in our generalized stochastic model.

Item limits or ”magic numbers” [2, 35] have previously been considered synonymous with discrete or slot-based accounts, occurring when some items must go unrepresented because other items have filled the available capacity. The present results show that a probabilistic item limit, i.e. a maximum on the average number of items successfully retrieved that is not

exceeded at any set size, can arise even when the probability of success for one item is independent of each other item. This holds true in the Poisson sampling model if we define success as obtaining one or more samples, but we have also shown it holds more generally, even in a continuous model, if we define success as exceeding a threshold level of precision in estimation.

Acknowledgments

We thank Máté Lengyel, Masud Husain and Wei Ji Ma for helpful discussions, and the researchers who publicly shared data that facilitated this study. This work was supported by the Wellcome Trust (grant no. 106926).

References

[1] S. J. Luck and E. K. Vogel. “The capacity of visual working memory for features and conjunctions”. *Nature* 390.6657 (1997), pp. 279–281.

[2] N. Cowan. “The magical number 4 in short-term memory: A reconsideration of mental storage capacity”. *Behavioral and brain sciences* 24.1 (2001), pp. 87–114.

[3] W. Zhang and S. J. Luck. “Discrete fixed-resolution representations in visual working memory”. *Nature* 453.7192 (2008), pp. 233–235.

[4] P. M. Bays, R. F. Catalao, and M. Husain. “The precision of visual working memory is set by allocation of a shared resource”. *Journal of Vision* 9.10 (2009), pp. 7–7.

[5] W. J. Ma, M. Husain, and P. M. Bays. “Changing concepts of working memory”. *Nature neuroscience* 17.3 (2014), p. 347.

[6] P. Wilken and W. J. Ma. “A detection theory account of change detection”. *Journal of Vision* 4.12 (2004), pp. 1120–1135.

[7] P. M. Bays and M. Husain. “Dynamic shifts of limited working memory resources in human vision”. *Science* 321.5890 (2008), pp. 851–854.

[8] D. Fougnie, J. W. Suchow, and G. A. Alvarez. “Variability in the quality of visual working memory”. *Nature communications* 3 (2012), p. 1229.

[9] R. van den Berg et al. “Variability in encoding precision accounts for visual short-term memory limitations”. *Proceedings of the National Academy of Sciences* 109.22 (2012), pp. 8780–8785.

[10] P. M. Bays. “Noise in neural populations accounts for errors in working memory”. *Journal of Neuroscience* 34.10 (2014), pp. 3632–3645.

[11] A. Pouget, P. Dayan, and R. Zemel. “Information processing with population codes”. *Nature Reviews Neuroscience* 1.2 (2000), p. 125.

[12] M. L. Shaw. “Identifying attentional and decision-making components in information processing”. *Attention and performance VIII* 8 (1980), pp. 277–295.

[13] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner. “Neuronal population coding of movement direction”. *Science* 233.4771 (1986), pp. 1416–1419.

[14] J. Palmer. “Attentional limits on the perception and memory of visual information”. *Journal of Experimental Psychology: Human Perception and Performance* 16.2 (1990), p. 332.

[15] D. K. Sewell, S. D. Lilburn, and P. L. Smith. “An information capacity limitation of visual short-term memory”. *Journal of Experimental Psychology: Human Perception and Performance* 40.6 (2014), p. 2214.

[16] P. L. Smith. “The Poisson shot noise model of visual short-term memory and choice response time: Normalized coding by neural population size”. *Journal of Mathematical Psychology* 66 (2015), pp. 41–52.

[17] R. van den Berg, E. Awh, and W. J. Ma. “Factorial comparison of working memory models.” *Psychological review* 121.1 (2014), p. 124.

[18] K. C. Adam, E. K. Vogel, and E. Awh. “Clear evidence for item limits in visual working memory”. *Cognitive psychology* 97 (2017), pp. 79–97.

[19] P. M. Bays et al. “Temporal dynamics of encoding, storage, and reallocation of visual working memory”. *Journal of vision* 11.10 (2011), pp. 6–6.

[20] A. H. Yoo et al. “Strategic allocation of working memory resource”. *Scientific reports* 8.1 (2018), p. 16162.

[21] S. M. Emrich, H. A. Lockhart, and N. Al-Aidroos. “Attention mediates the flexible allocation of visual working memory resources.” *Journal of Experimental Psychology: Human Perception and Performance* 43.7 (2017), p. 1454.

[22] Z. Klyszejko, M. Rahmati, and C. E. Curtis. “Attentional priority determines working memory precision”. *Vision research* 105 (2014), pp. 70–76.

[23] K. H. Britten et al. “The analysis of visual motion: a comparison of neuronal and psychophysical performance”. *Journal of Neuroscience* 12.12 (1992), pp. 4745–4765.

[24] E. Zohary, M. N. Shadlen, and W. T. Newsome. “Correlated neuronal discharge rate and its implications for psychophysical performance”. *Nature* 370.6485 (1994), p. 140.

[25] B. B. Averbeck, P. E. Latham, and A. Pouget. “Neural correlations, population coding and computation”. *Nature reviews neuroscience* 7.5 (2006), p. 358.

[26] A. S. Ecker et al. “The effect of noise correlations in populations of diversely tuned neurons”. *Journal of Neuroscience* 31.40 (2011), pp. 14272–14283.

[27] R. Moreno-Bote et al. “Information-limiting correlations”. *Nature neuroscience* 17.10 (2014), p. 1410.

[28] J. M. Beck et al. “Not noisy, just wrong: the role of suboptimal inference in behavioral variability”. *Neuron* 74.1 (2012), pp. 30–39.

[29] R. Taylor and P. M. Bays. “Efficient coding in visual working memory accounts for stimulus-specific variations in recall”. *Journal of Neuroscience* 38.32 (2018), pp. 7132–7142.

[30] X.-X. Wei and A. A. Stocker. “A Bayesian observer model constrained by efficient coding can explain’anti-Bayesian’percepts”. *Nature neuroscience* 18.10 (2015), p. 1509.

[31] D. Ganguli and E. P. Simoncelli. “Efficient sensory encoding and Bayesian inference with heterogeneous neural populations”. *Neural computation* 26.10 (2014), pp. 2103–2134.

[32] P. M. Bays. “Correspondence between population coding and psychophysical scaling models of working memory”. *BioRxiv* (2019), p. 699884.

[33] H. Sompolinsky et al. “Population coding in neuronal systems with correlated noise”. *Physical Review E* 64.5 (2001), p. 051904.

[34] S. Schneegans and P. M. Bays. “Neural architecture for feature binding in visual working memory”. *Journal of Neuroscience* 37.14 (2017), pp. 3913–3925.

[35] G. A. Miller. “The magical number seven, plus or minus two: Some limits on our capacity for processing information.” *Psychological review* 101.2 (1994), p. 343.