Vol 453 | 8 May 2008 | doi:10.1038/nature06860

Discrete fixed-resolution representations in visual working memory

Weiwei Zhang[1,2] & Steven J. Luck[2]

Limits on the storage capacity of working memory significantly affect cognitive abilities in a wide range of domains[1], but the nature of these capacity limits has been elusive[2]. Some researchers have proposed that working memory stores a limited set of discrete, fixed-resolution representations[3], whereas others have proposed that working memory consists of a pool of resources that can be allocated flexibly to provide either a small number of high-resolution representations or a large number of low-resolution representations[4]. Here we resolve this controversy by providing independent measures of capacity and resolution. We show that, when presented with more than a few simple objects, human observers store a high-resolution representation of a subset of the objects and retain no information about the others. Memory resolution varied over a narrow range that cannot be explained in terms of a general resource pool but can be well explained by a small set of discrete, fixed-resolution representations.

To separately measure the number of items stored in working memory and the precision of each representation, we used a short-term recall paradigm[5,6] in which subjects report the remembered colour of a probed item by clicking on a colour wheel (Fig. 1a). If the probed item has been stored in working memory, the recalled value will tend to be near the original colour. If the probed item has not been stored, then the observer will have no information about the colour, and the responses should be random. These two types of trials are mixed together in the data (Fig. 1b), but the components can be recovered via standard estimation methods. This produces one parameter (P_m) representing the probability that the probed item was present in memory at the time of the probe and another parameter (s.d.) representing the precision of the representation when the cued item was present in memory.

Experiment 1 ($N = 8$) tested this model using set sizes of 3 or 6 coloured squares (Fig. 1c). s.d. did not vary significantly across set sizes ($F < 1$), whereas P_m was approximately twice as great at set size 3 as at set size 6 ($F(1,7) = 761.26$, $P < 0.001$). Our simple fixed-resolution model provided an excellent quantitative fit to the data, whereas a model in which all items are encoded could not fit the data (see Supplementary Notes). This result rules out the entire class of working memory models in which all items are stored but with a resolution or noise level that depends on the number of items in memory[5]. Control experiments demonstrated that these results cannot be explained by a lack of time to encode the items or by a lack of sensitivity, and additional analyses demonstrated that the observers remembered continuous colour values rather than colour categories (see Supplementary Notes).

These results demonstrate that observers store a small number of representations with good precision. However, it is possible that performance is influenced both by a limited number of 'storage slots' and a limited pool of resources[7]. As an analogy, consider three cups (the slots) and a bottle of juice (the resource). It would be impossible to serve juice to more than three people at a time, but it would be possible to pour most of the juice into a single cup, leaving only a few drops for the other two cups. Thus, allocating most of the resources to a single representation could increase the precision of that representation, leaving 'only a few drops' of resources for the other representations, which would then be highly imprecise. We call this the 'slots+resources' model.

The storage of information in visual working memory could instead be an all-or-none process that either creates a representation of a given precision or creates no representation at all. This would be analogous to a limited set of prepackaged juice boxes of a fixed size. The juice boxes are still a type of resource, but one that is highly constrained by the small number and fixed size of each box. That is, if three juice boxes are available, an individual could be given 0, 1, 2 or 3 boxes. Similarly, if three memory slots are available, all three could be used to represent a single object. If each representation stores an independent sample of the stimulus, and observers simply report the average of the three representations at the time of test, this will lead to an increase in the precision of the report. We call this the 'slots+averaging' model. Note that storing a single object in multiple

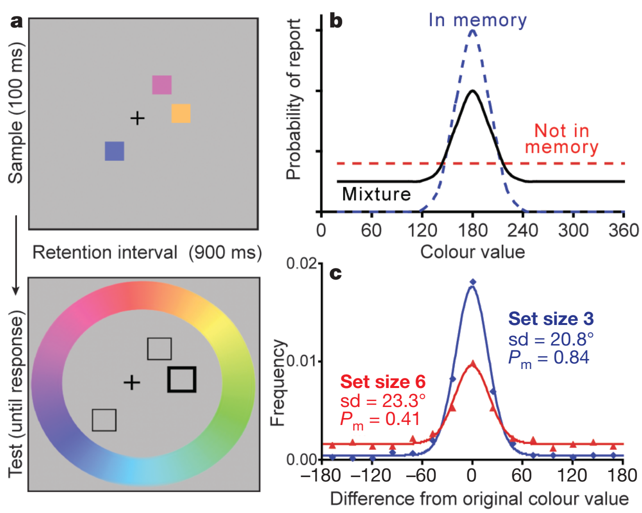


Figure 1 | Experimental approach and results of experiment 1. a, Colour recall task. **b,** Mixture model of performance, showing the probability of reporting each colour value given a sample colour at 180°. When the probed item is present in memory, the reported colour tends to be near the original colour (blue broken line). When the probed item is not present in memory, the observer is equally likely to report any colour value (red broken line). When collapsed across trials, the data comprise a mixture of these two trial types (solid line), weighted by the probability that the probed item was stored in memory. **c,** Results of experiment 1 ($N = 8$). P_m and s.d. are defined in the text.

[1]Department of Psychology, University of Iowa, Iowa City, Iowa 52242, USA. [2]Center for Mind & Brain, University of California, Davis, California 95618, USA.

slots would be a rational strategy, and this sort of averaging is common in models of perception[8–10].

For both the slots+resources and slots+averaging models, s.d. will be improved when the set size is reduced below the number of available slots. Moreover, both models predict that this improvement will follow a square root function (see Supplementary Notes). This is exactly what was observed in experiment 2 (Fig. 2), in which observers ($N = 8$) were presented with 1, 2, 3 or 6 objects: s.d. increased as the set size increased from 1 to 3 but then remained constant as the set size increased to 6. In contrast, P_m declined very slowly as set size increased from 1 to 3 and then decreased suddenly at set size 6. This pattern of results can be explained quantitatively by both the slots+resources model (adjusted $r^2 = 0.96$) and the slots+averaging model (adjusted $r^2 = 0.99$) (see Fig. 2 and Supplementary Notes), but it differs significantly from the predictions of a pure resource model ($P < 0.001$, χ^2 test).

The slots+resources and slots+averaging models make different predictions about the range over which precision can vary. Specifically, the slots+resources model posits that the majority of resources can be devoted to one representation (leading to a very small s.d.), leaving 'only a few drops' of resources for other representations (leading to a very large s.d.). In contrast, the slots+averaging model posits that the observed s.d. is never worse than the s.d. of a single slot and is never better than the s.d. for a single slot divided by the square root of the number of slots. To distinguish between these models, experiment 3 ($N = 22$) used a line in the sample array to cue one of four coloured squares (Fig. 3a). The cued square was probed on 70% of the trials, and each uncued square was probed on 10% of trials. Neutral trials were also included, in which all four locations were cued. The cue was simultaneous with the sample array so that it would not influence perceptual processing[11], and the duration of the sample array was increased to 300 ms to provide adequate time for resource allocation[10,12].

The slots+resources model predicts that observers will devote the lion's share of resources to the cued item, leading to a large difference in s.d. between valid, neutral and invalid trials, but only a small difference in P_m. In contrast, the slots+averaging model predicts that observers will devote most of their slots to the cued location, which would lead to a large difference in P_m between valid and invalid trials. This should also lead to a somewhat smaller s.d. on valid trials than on neutral trials because of the benefits of averaging. However, this should lead to no difference in s.d. between neutral and invalid trials, because a given item receives either 0 or 1 slots on both neutral and invalid trials.

We found that P_m was substantially greater on valid trials than on invalid trials ($F(1,21) = 203.87$, $P < 0.001$; Fig. 3a), demonstrating that the observers attempted to maximize performance for the cued item by devoting more slots to it. s.d. was slightly but significantly smaller on valid trials than on neutral trials ($F(1,21) = 13.49$, $P < 0.001$), and the magnitude of this difference was within the small range that can result from averaging slots. In addition, s.d. was virtually identical on neutral and invalid trials ($F < 1$), indicating that the improvement in s.d. on valid trials was not achieved by taking resources away from the uncued items. Thus, despite the fact that the cued item was seven times more likely to be probed than each uncued item, s.d. was only slightly improved for the cued item (compared to the neutral trials) and s.d. was not reduced for the uncued items (compared with the neutral trials). It does not appear to be possible to provide a representation with 'only a few drops' of resources and thereby produce an imprecise representation.

Computational neuroscience theories suggest that an all-or-none, fixed-resolution encoding process may be required to create durable representations that can survive new sensory inputs[13,14]. To assess the encoding process, experiment 4 ($N = 8$) used a masking manipulation that emulates the masking effects of eye movements in natural vision. Specifically, we presented masks at the locations of the coloured squares either 110 or 340 ms after the onset of the squares (Fig. 3b). At these intervals, masks interfere with working memory encoding but not with perceptual processing[15]. If working memory representations gradually become more precise over time, then presenting a mask array at an early time point could potentially reveal the existence of low-precision representations. If, however, the process of creating durable memory representations that can survive new visual inputs involves an all-or-none step, as suggested by studies of the 'attentional blink' phenomenon[16], then the masks will only influence P_m. We observed that decreasing the masking interval produced a large decline in P_m ($F(1,7) = 47.70$, $P < 0.001$) but no change in s.d. ($F < 1$). Thus, the creation of working memory representations that can survive new inputs involves an all-or-none step for simple objects (although it is possible that some gradual accumulation of information occurs before this step and is available in the absence of masking).

To demonstrate that the present results can generalize to other stimulus dimensions, we repeated experiments 2 and 3 with shapes rather than colours. We used shapes defined by Fourier descriptors[17], which vary along continuous quantitative dimensions. The results were largely identical to the results obtained for colour, with approximately the same P_m for these shapes as for the simple colours in experiment 2 (see Supplementary Notes). Most notably, s.d. did not increase as the set size increased from three to six items and was virtually identical for neutral and invalid trials. Because the Fourier descriptor method provides a mathematically[17], perceptually[18,19] and neurally[20] meaningful way to describe shapes of any complexity, this analytic approach could be used to determine whether the present pattern of results would be obtained with more complex objects. Object complexity can have a large impact on performance in change detection tasks[7], but this may reflect greater sample-test similarity for complex objects[21] or the need to store each

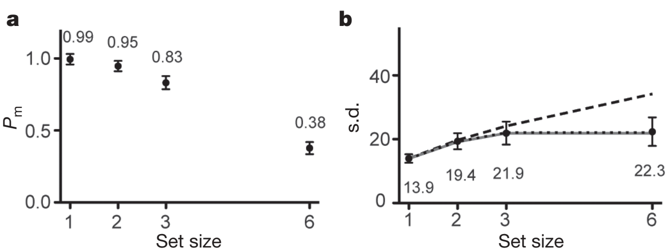


Figure 2 | P_m and s.d. results from experiment 2 ($N = 8$). **a**, P_m; **b**, s.d.; the numbers within the panels provide the means. In **b**, the lines show the s.d. predictions of a pure resource model (black dashed line), the slots+averaging model (grey solid line), and the slots+resources model (black dotted line). Error bars show within-subjects 95% confidence intervals[26].

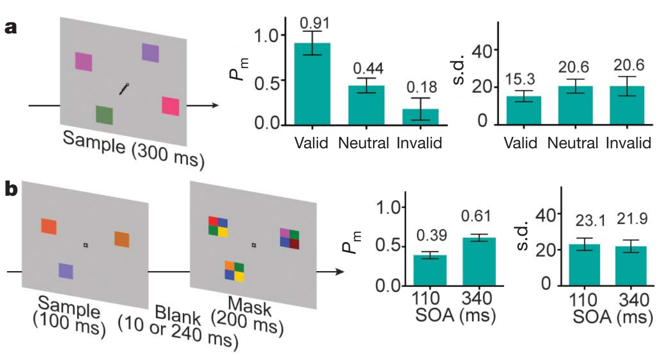


Figure 3 | Stimuli and results from experiments 3 and 4. **a**, Experiment 3 ($N = 22$), which included valid, neutral and invalid trials. The cue appeared simultaneously with the sample array. **b**, Experiment 4 ($N = 8$), in which a mask array followed the sample array with a stimulus onset asynchrony (SOA) of 110 or 340 ms. A colour wheel and probe array appeared at the end of the trial, 900 ms after sample offset. Error bars show within-subjects 95% confidence intervals[26], and numbers within the panels provide the means.

part of a multipart object in a separate slot[22,23]. Alternatively, complex objects may require some kind of limited resource that is not needed for the simple objects studied here.

Together, the present experiments resolve an issue that has been debated for decades[4,5,24,25], showing that a model with a small set of discrete, fixed-resolution representations can provide a quantitative account of memory performance across a broad range of experimental manipulations. This model does not completely eliminate the concept of resources, because the slots themselves are a type of resource. However, the slots+averaging model defines exactly what the resource is and describes strict limits on how flexibly this resource can be allocated.

METHODS SUMMARY

The stimuli and task are shown in Fig. 1a. Subjects viewed a sample array and then, following a brief delay, reported the colour of one item from this array (indicated by a thick outlined box) by clicking on a colour wheel. There is good agreement between this procedure and the more commonly used change-detection procedure (see Supplementary Notes).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 November 2007; accepted 14 February 2008.
Published online 2 April 2008.

1. Baddeley, A. D. *Working Memory* (Clarendon, Oxford, 1986).
2. Miyake, A. & Shah, P. (eds) *Models of Working Memory* (Cambridge Univ. Press, Cambridge, UK, 1999).
3. Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390,** 279–281 (1997).
4. Frick, R. W. Issues of representation and limited capacity in the visuospatial sketchpad. *Br. J. Psychol.* **79,** 289–308 (1988).
5. Wilken, P. & Ma, W. J. A detection theory account of change detection. *J. Vis.* **4,** 1120–1135 (2004).
6. Prinzmetal, W., Amiri, H., Allen, K. & Edwards, T. Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *J. Exp. Psychol. Hum. Percept. Perform.* **24,** 261–282 (1998).
7. Alvarez, G. A. & Cavanagh, P. The capacity of visual short-term memory is set both by information load and by number of objects. *Psychol. Sci.* **15,** 106–111 (2004).
8. Palmer, J. Attentional limits on the perception and memory of visual information. *J. Exp. Psychol. Hum. Percept. Perform.* **16,** 332–350 (1990).
9. Newsome, W. T. *et al.* in *The Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 401–414 (MIT Press, Cambridge, Massachusetts, 1995).
10. Bonnel, A. M. & Miller, J. Attentional effects on concurrent psychophysical discriminations: Investigations of a sample-size model. *Percept. Psychophys.* **55,** 162–179 (1994).
11. Vogel, E. K., Woodman, G. F. & Luck, S. J. Pushing around the locus of selection: Evidence for the flexible-selection hypothesis. *J. Cogn. Neurosci.* **17,** 1907–1922 (2005).
12. Luck, S. J., Hillyard, S. A., Mouloua, M. & Hawkins, H. L. Mechanisms of visual-spatial attention: Resource allocation or uncertainty reduction? *J. Exp. Psychol. Hum. Percept. Perform.* **22,** 725–737 (1996).
13. Wang, X.-J. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* **24,** 455–463 (2001).
14. Raffone, A. & Wolters, G. A cortical mechanism for binding in visual working memory. *J. Cogn. Neurosci.* **13,** 766–785 (2001).
15. Vogel, E. K., Woodman, G. F. & Luck, S. J. The time course of consolidation in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* **32,** 1436–1451 (2006).
16. Sergent, C. & Dehaene, S. Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychol. Sci.* **15,** 720–728 (2004).
17. Zahn, C. T. & Roskies, R. Z. Fourier descriptors for plane closed curves. *IEEE Trans. Comput. C* **21,** 269–281 (1972).
18. Alter, I. & Schwartz, E. L. Psychophysical studies of shape with Fourier descriptor stimuli. *Perception* **17,** 191–202 (1988).
19. Cortese, J. M. & Dyre, B. P. Perceptual similarity of shapes generated from Fourier descriptors. *J. Exp. Psychol. Hum. Percept. Perform.* **22,** 133–143 (1996).
20. Schwartz, E. L., Desimone, R., Albright, T. D. & Gross, C. G. Shape recognition and inferior temporal neurons. *Proc. Natl Acad. Sci. USA* **80,** 5776–5778 (1983).
21. Awh, E., Barton, B. & Vogel, E. K. Visual working memory represents a fixed number of items regardless of complexity. *Psychol. Sci.* **18,** 622–628 (2007).
22. Xu, Y. Limitations of object-based feature encoding in visual short-term memory. *J. Exp. Psychol. Hum. Percept. Perform.* **28,** 458–468 (2002).
23. Sakai, K. & Inui, T. A feature-segmentation model of short-term visual memory. *Perception* **31,** 579–590 (2002).
24. Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* **63,** 81–97 (1956).
25. Simon, H. A. How big is a chunk? *Science* **183,** 482–488 (1974).
26. Cousineau, D. Confidence intervals in within-subjects designs: A simpler solution to Loftus and Masson's method. *Tutor. Quant. Methods Psychol.* **1,** 42–45 (2007).
27. Myung, I. J. Tutorial on maximum likelihood estimation. *J. Math. Psychol.* **47,** 90–100 (2003).
28. Fisher, N. I. *Statistical Analysis of Circular Data* (Cambridge Univ. Press, New York, 1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was made possible by grants from NIMH to S.J.L.

Author Contributions W.Z. and S.J.L. together developed the general conceptualization of this study and wrote the paper. W.Z. designed and implemented the experimental and analytical approach.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.Z. (wwzhang@ucdavis.edu).

METHODS

Subjects. Eight subjects between 18 and 35 years old participated in each colour memory experiment except the cuing experiment, in which 22 subjects participated owing to the low probability of the invalid trials. Subjects provided informed consent and received course credit or monetary compensation. All reported having normal colour vision and normal or corrected-to-normal visual acuity.

Experiments. The stimuli in all experiments were presented on a CRT monitor with a grey background ($15.5\,\mathrm{cd\,m^{-2}}$) at a viewing distance of 57 cm. The monitor was calibrated with a Tektronix J17 LumaColour colorimeter. Each coloured square in the sample array subtended $2 \times 2^\circ$ of visual angle. Each square was centred on an invisible circle with a radius of 4.5°. The positions were randomly chosen from a set of eight locations equally spaced along the circle. The colour wheel was 2.2° thick and was centred on the monitor with a radius of 8.2°. It consisted of 180 colour values that were evenly distributed along a circle in the CIE $L^*a^*b^*$ colour space. This circle was centred in the colour space at ($L = 70$, $a = 20$, $b = 38$) with a radius of 60. Its centre was chosen to maximize its radius and therefore the discriminability of the colours. All colours had equal luminance and varied mainly in hue and slightly in saturation. The sample array colours were randomly selected from this set of colours. The colour wheel was presented at a random rotation on each trial to minimize contributions from spatial memory.

In the basic paradigm (experiments 1 and 2), each trial consisted of a 100-ms sample array followed by a 900-ms blank delay period and then a probe display that remained present until a response was made (Fig. 1a). The probe display contained the colour wheel and an outlined square at the location of each item from the sample array. One of these squares was thicker (0.20°) than the others (0.04°), which cued the subject to recall the colour of the corresponding item from the sample array by clicking the appropriate colour on the colour wheel with the computer mouse. Accuracy was stressed, and the responses were not timed. Except as noted below, 150 trials were tested in each experimental condition (for example, each set size). The different trial types (for example, different set sizes) were presented in an unpredictable order in each experiment.

In the control experiment that involved varying the level of perceptual noise (Supplementary Fig. 1), the set size was held constant at three items and the duration of the sample array was reduced to 30 ms to ensure that the masks would be effective. Each coloured square in the sample array was covered with a set of either 75 or 150 simultaneously presented coloured dots, randomly distributed over a circular region with a diameter of 4.4° that was centred on the coloured square. Each dot subtended $0.2 \times 0.2^\circ$ of visual angle and was drawn in a colour that was randomly sampled from the set of 180 colour values used for the coloured squares.

In the cuing experiment (Fig. 3a), the sample display contained a 1.6° cue line extending from fixation towards one of the four coloured squares (on valid and invalid trials) or four lines extending towards all four squares (on neutral trials). The duration of the sample display was increased to 300 ms in this experiment to provide the observers sufficient time to shift attention to the cued item; the interval between sample onset and probe onset remained 900 ms. Each observer received 350 valid trials, 150 invalid trials (50 per uncued location) and 150 neutral trials. These trial types were randomly intermixed.

In the backward masking experiment (Fig. 3b), the sample array always contained three items. An array of masks was presented 110 or 340 ms after the onset of the sample array, with a duration of 200 ms. The interval between sample offset and probe onset remained constant at 900 ms. Each mask consisted of a 2×2 arrangement of coloured squares, each of which measured $0.55 \times 0.55^\circ$. Each mask was centred at the location of one of the three items in the sample display.

Data analysis. The data from a given observer in the colour experiments consisted of a set of distances between the reported colour value and the original colour value in each condition, which reflects the degree of error in the reported colour. Histograms of these error values were used to visualize the distribution of responses (as in Fig. 1c). Maximum likelihood estimation[27] was used to decompose the data from each subject in each condition into three parameters that represent a mixture of a uniform distribution of errors (for trials on which the probed item was not encoded in memory) and a von Mises distribution of errors (for trials on which the probed item was encoded). The von Mises distribution is the circular analogue of the gaussian distribution and was used because the tested colour space was circular[28]. The uniform distribution was represented by a single parameter, P_m, which is the probability that the probed item was present in memory at the time of the probe (which is inversely related to the height of the uniform distribution). The von Mises distribution was represented by two parameters, its mean (μ) and its standard deviation (s.d.). μ reflects any systematic shift of the distribution away from the original colour value. No systematic shifts were expected or observed in any of the present experiments, so this parameter will not be considered further. s.d. reflects the width of the distribution of errors on trials when the probed item was encoded in memory, which in turn reflects the precision or resolution of the memory representation.

The slots+averaging model was fitted to the estimated P_m and s.d. parameters in the experiment in which set sizes 1, 2, 3 and 6 were tested. We computed the total number of slots (K_i) by multiplying P_m by the set size (using the data from set size 3). We then assumed that the slots were randomly distributed among the available items in the sample array, allowing multiple slots to be assigned to a given object if the set size was lower than the number of slots. The s.d. at set size 3 was used to estimate the precision of a single slot. The s.d. from a set of N samples is equal to the s.d. from a single sample divided by the square root of N (see Supplementary Notes). Thus, by knowing the s.d. of a single slot and the average number of slots assigned to the probed item in a given condition, it is possible to predict the s.d. for that condition.

In the slots+resources model, the s.d. at set size 1 was used to estimate the maximum precision when all resources are devoted to a single object. For modelling the data from larger set sizes, the s.d. simply increases as a function of the square root of the number of objects being represented, up to the number of slots (which is estimated as in the slots+averaging model). For simple manipulations of set size, the predictions of the slots+resources model are equivalent to those of the slots+averaging model except that the s.d. values are estimated on the basis of the data at set size 1 rather than the data at set size 3.

Supplementary Notes

Relationship Between Slots, Resources, and SD

The slots+averaging model supposes that observers may store multiple independent samples of a colour in separate slots, reporting the average of the stored colour values at the time of the memory probe. The SD of the average of N samples is well known to be equal to the SD of the individual samples divided by sqrt(N)[1]. This fact was used to make quantitative predictions from the slots+averaging model.

Resource models are analogous to slot models with an infinite number of slots. Consequently, quantitative resource models also lead to a square-root relationship between the quantity of resources devoted to a representation and the SD of that representation[1-3]. The essential difference between slot and resource models is the granularity of the resource, which is infinitely divisible in resource models but is organized into a few large chunks in slot models.

Adequacy of Parameter Estimations

As evidence that our mixture model with three parameters provides an adequate description of the data, we computed the adjusted r^2 statistic (which reflects the proportion of variance explained by the model) and the χ^2 statistic on the basis of histograms of the data with 15 bins, each 24° wide. The Kolmogorov-Smirnov (K-S) statistic was also computed to test whether the observed data differ significantly from the model. These statistics were computed for each individual subject and also for the data aggregated over subjects. It is more difficult to obtain a good fit when P_m is low and when a given condition has relatively few trials per subject, and the group data sets were particularly useful for demonstrating goodness of fit under these conditions. We also conducted 1000 Monte Carlo simulations of the data that would be expected for individual observers in each condition to determine how the adjusted r^2 values would be expected to vary given the number of trials and the observed P_m values.

Supplementary Table 1 shows the adjusted r^2 values for each major experiment described in the main text, including the values computed from the group data, the mean of the single-subject data, and the mean of the simulated single-subject data. In the single-subject data, the worst fits explained an average of over 75% of the variance, and the best fits explained an average of over 95% of the variance. The variations in goodness of fit corresponded well with variations observed in the simulations, indicating that the variations in goodness of fit are a consequence of variations in the number of trials and P_m levels rather than reflecting systematic deviations of the data from the model. For example, although the observed adjusted r^2 values fell from approximately 0.95 at set sizes 1–3 to approximately 0.87 at set size 6, a similar drop was observed in the simulations, presumably because of an increase in the proportion of trials on which subjects did not recall the probed colour and therefore responded randomly. Thus, the reduction in goodness of fit was an inevitable consequence of the nature of the underlying memory representations in the context of a finite data set. In addition, the model explained over 95% of the variance in all conditions when the data were aggregated across the group of subjects. Moreover, the K-S and χ^2 analyses indicated that the observed data were not significantly different from the model in any subject or group of subjects for any condition. Thus, this simple 3-parameter model provides an excellent quantitative fit to the data across all conditions of all experiments.

For Experiments 1 and 2, we also tested the adequacy of a simple resource model containing only a von Mises distribution, which is equivalent to holding P_m constant at 1.0 when estimating the μ and SD parameters. The adjusted r^2 values were negative for all set sizes in both experiments, with the exception of set size 1 in Experiment 2. A negative value means that the error variance is larger than what would be obtained by a model in which the mean of the sample was the only parameter, and it indicates that the model does not adequately fit the data. In addition, the data deviated significantly from the model according to the K-S and χ^2 tests for all set sizes in both experiments ($p < .05$ or better). Thus, the data are quantitatively inconsistent with a model in which all items are encoded and only the precision of the representations varies as a function of set size[4].

Control Experiments

To demonstrate that the effect of set size on P_m did not reflect a lack of sufficient time to encode the items at set size 6, we conducted a control experiment comparing sample durations of 100 and 500 ms at set size 6. We found no significant effect of duration on either SD or P_e (both $Fs < 1$). To demonstrate that our methods are sufficiently sensitive to detect changes in SD if they are present, we conducted an additional control experiment in which the quality of the perceptual representations was manipulated by adding varying numbers of coloured "noise" dots to a set of 3 coloured squares (see Supplementary Figure 1). The noise degraded the perceptual representation of the colours, and this reduced precision was necessarily propagated to the memory representations. Increasing the noise increased the SD ($F(1,7)= 7.78$, $p<0.03$) but did not influence P_m ($F < 1$). Thus, our methods are sufficiently sensitive to detect modest changes in precision.

Colour Categories

Although it is not central to this study, our methods implicitly assume that observers store a representation of the continuous colour values. However, it is possible that they instead convert the continuous colour values in the sample array into categorical representations (e.g., prototypes of red, green, blue, etc.). If this were true, then much of the distribution of responses would reflect the difference between the actual colour of the probed item and the nearest prototypical colour value. To assess this possibility, we pooled the data from four experiments that included trials with a set size of one item (the procedure for these trials was identical across experiments), yielding a total of 50 observers. (Pooling across observers is well justified because colour categories are highly consistent across individuals from a restricted age range and cultural group[5].) We then plotted the distribution of reported colours as a function of the actual colours for this pool of observers (Supplementary Figure 2a). If observers represent the actual colour (plus noise), then this should yield a straight line. If observers instead represent a given sample colour as the nearest colour category value, then this function should look like a staircase, in which variations in the actual colour within a given range lead to no change in the reported colour, with a sudden change in reported colour when the actual colour crosses the category boundary. An example of this is displayed in Supplementary Figure 2b, which shows the results of a simulation of categorical memory with 7 colour categories. The results shown in Supplementary Figure 2a clearly follow a straight line, and a least-squares analysis showed that a straight line accounts for 97% of the variance in reported colour. There was no sign of staircase-like horizontal bands in these data. Thus, observers appear to remember the actual colour rather than the nearest colour prototype.

Comparison of Colour Recall with Colour Change Detection

The colour recall task used in this study differs from the more common change detection task that has been used widely to study visual working memory in behavioural[6-8], ERP[9,10], and neuroimaging[11,12] studies. To determine whether these two tasks are measuring the same aspects of working memory, we conducted an experiment (N=14) in which the two tasks were randomly intermixed. Each trial began with a 100-ms presentation of a sample array containing three items. This was followed after a 900-ms delay by either a colour recall test display or a change-detection test display.

The colour recall test display contained a probe and a colour wheel, just as in the other experiments in the present study, and observers responded by selecting a colour from the colour wheel. The change detection test display contained a single coloured square at one of the locations from the sample array, and observers responded by making a keypress response to indicate whether or not the test square was the same colour as the corresponding sample square. The test colour was the same on 50% of trials and differed by 180° in colour space on the other 50%. The colour recall and change detection trials were unpredictably intermixed, so observers necessarily encoded and maintained the colour information in the same way on both trial types.

The goal of this experiment was to determine whether the estimated number of items that observers store in memory (the storage capacity) is the same for the two tasks. The storage capacity for the colour recall task (K_i) was estimated by simply multiplying P_m by the set size. The storage capacity for the change detection task was estimated with the Cowan K equation[13]. This equation is based on a high-threshold model, but it should be approximately correct for the maximally large (180°) change magnitudes used in this experiment. As shown in Supplementary Figure 3, the two measures of storage capacity were strongly and significantly correlated across subjects ($r^2 = .572$, $p = .002$), with a slope near 1.0 and an intercept near 0.0. Although the agreement between these two procedures may vary depending on the decision requirements of the specific experiment, these results suggest that they are measuring fundamentally similar aspects of visual working memory capacity.

Extension to Shape-Defined Stimuli

Experiments 2 and 3 were repeated with shape stimuli (N = 8 and 14, respectively; see stimuli in Supplementary Figure 4 and results in Supplementary Figure 5). Shape was parameterized using the Fourier descriptor technique[14], in which the perimeter of an object is described by the sum of a set of sinusoidal components. That is, a function is created that represents the distance between the centre of the object and its perimeter as a function of polar angle, and this function is then decomposed into the sum of a set of sine wave components that vary in frequency, amplitude, and phase. Using this approach, we synthesized a family of objects consisting of two sinusoidal components, one with a frequency of 2 cycles per perimeter (cpp) and an amplitude of 0.5 and one with a frequency of 4 cpp and an amplitude of 0.5. The phase of the 2-cpp component was held constant at 0°, and the phase of the 4-cpp component varied between 0 and 360° in steps of 2° (providing a circular dimension that is analogous to the hue dimension used in the colour experiments). The result was a family of 180 gradually varying shapes, each subtending approximately 2° in visual angle.

The procedure for the shape stimuli was identical to that used for the colour stimuli, with three exceptions. First, the exposure duration was increased to 1000 ms to provide sufficient encoding time. Second, the circle of shapes at the time of response contained 30 discrete shapes rather than a visually continuous ring of 180 colour values (see Supplementary Figure 4). These shapes were sampled from the family of 180 shapes, with 12° of phase difference between adjacent shapes and a randomly chosen starting phase on each trial. Subjects were instructed to indicate the remembered shape of the probed item by clicking on the corresponding position within the circle of shapes, interpolating between the exemplars in the circle if necessary to accurately report the shape (because the actual shape may lie between two of the sample shapes in a given display). Third, a familiarization block with 60 trials was run before the memory task, in which six shapes were presented simultaneously with the circle of shapes. One of the six shapes was cued and subjects matched the cued shape to the corresponding shape on the circle of shapes. This provided the observers with an opportunity to practice choosing interpolated locations along the circle of test shapes, and the experimenter verified that they did so. The data from the shape experiments were analyzed in exactly the same manner as the data from the corresponding colour experiments.

Overall P_m values for shape were similar to those obtained for colour. The SD was somewhat larger for shape than for colour, but there is no reason why the precision should have the same numeric value across dimensions, especially given that the test display contained 30 values for shape and 180 values for colour. When the set size was manipulated as in Experiment 2, P_m fell slightly as the set size increased from 1 to 3 and then fell dramatically as the set size increased from 3 to 6, producing a significant overall effect of set size ($F(3,21) = 51.76$, $p < 0.001$). SD increased significantly as the set size increased between 1 and 3 items ($F(2,14) = 13.15$, $p < 0.001$) but did not increase as the set size increased from 3 to 6 items ($F < 1$). This is the same pattern of results obtained for colour in Experiment 2, and it was well fit by both the slots+resources and slots+averaging models but not by the simple resource model.

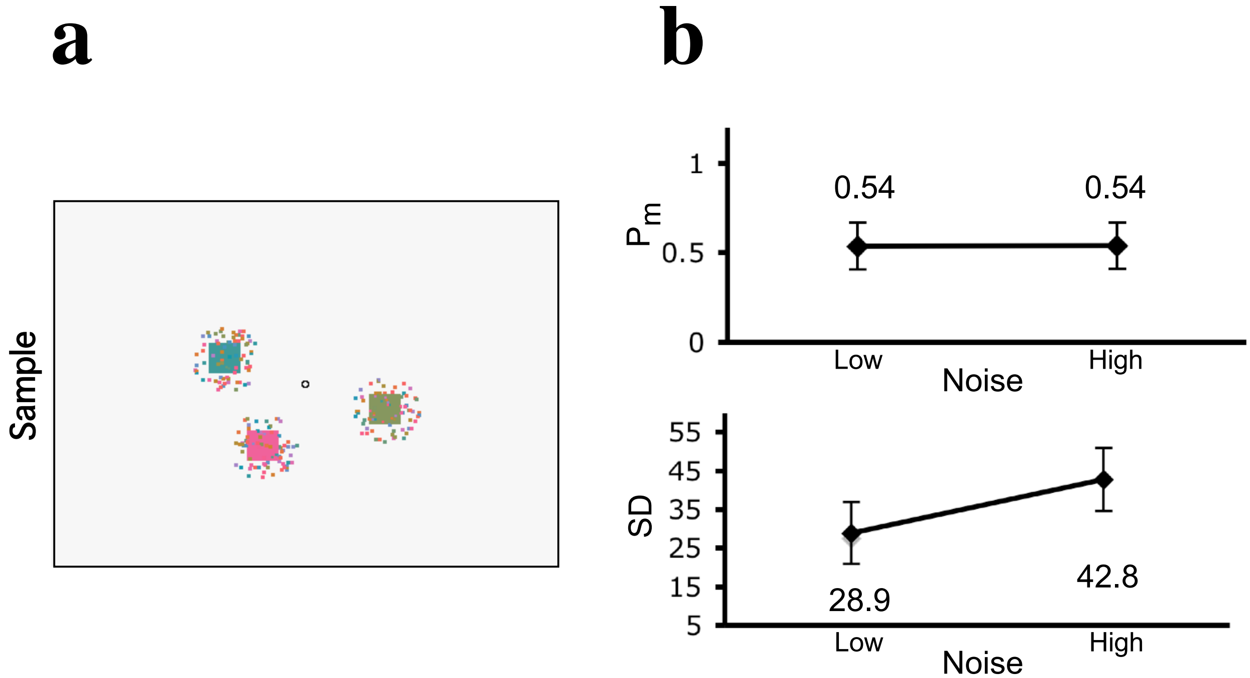
When a cue in the initial sample array was used to direct attention to a single item, as in Experiment 3, P_m was high for valid trials, substantially lower on neutral trials, and very low on invalid trials ($F(2,26)=15.23$, $p<0.001$). This demonstrates that the cue was highly effective in motivating the observers to give priority to the cued shape. However, cuing produced only a modest difference in SD between valid and neutral trials ($F(1,13)=12.54, p<0.01$) and no difference between neutral and invalid trials ($F < 1$). This pattern is identical to the pattern observed for colour in Experiment 3. The size of the improvement in SD on valid trials compared to neutral trials was within the range that would be expected if the observers allocated multiple slots to the cued item on valid trials and allocated a single slot to each item on neutral trials. Most importantly, the lack of an increase in SD on invalid trials compared to neutral trials indicates that focusing attention onto the cued item did not result in reduced precision for the uncued items. That is, uncued items were relatively unlikely to be stored in memory, but when they were stored they were represented with the same precision as on neutral trials. Thus, it does not seem possible to allocate "only a few drops" of resources to a shape representation in working memory.

Supplementary References

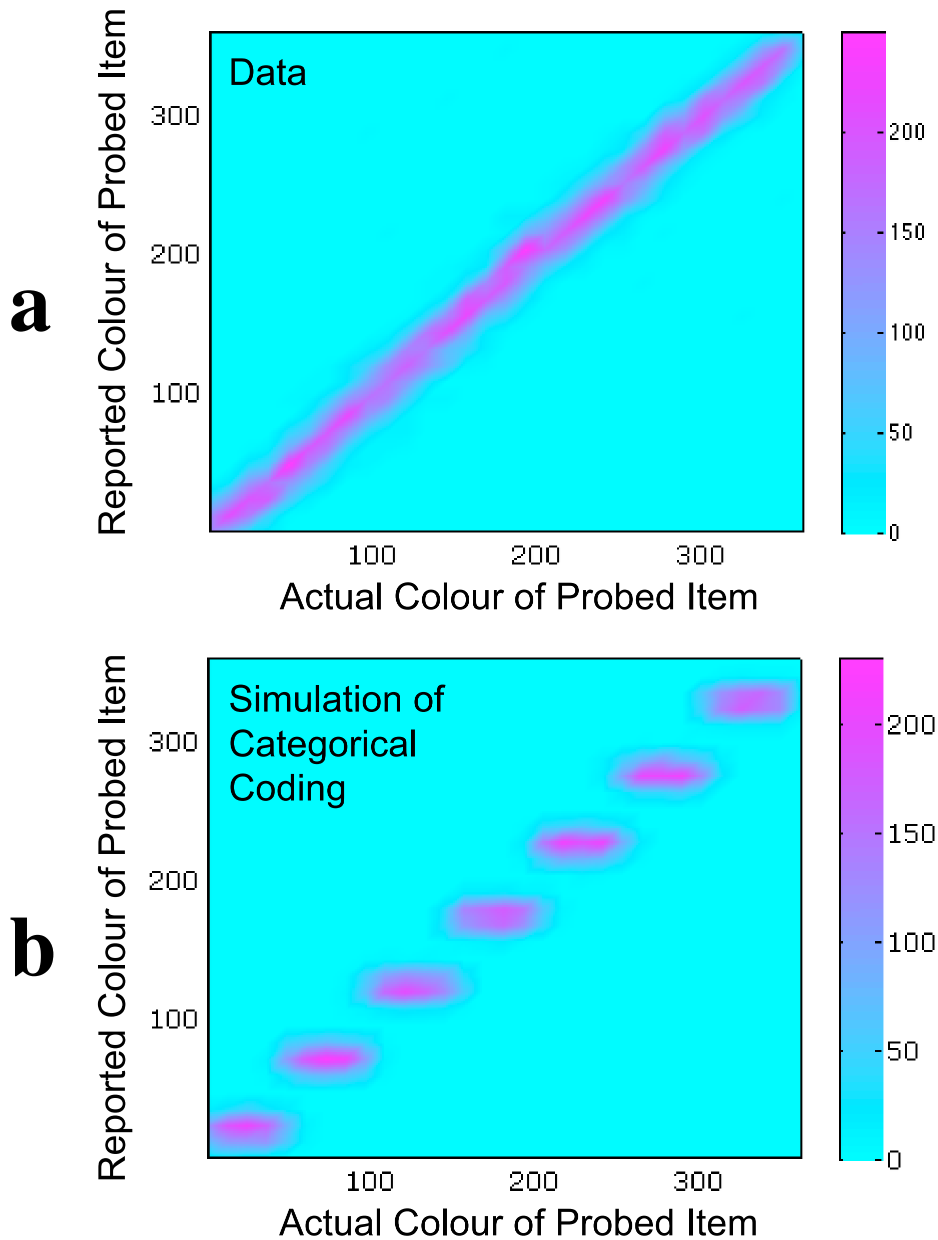
1. Palmer, J. Attentional limits on the perception and memory of visual information. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 332-350 (1990).
2. Shaw, M. L. in *Attention and performance*, edited by R. S. Nickerson (L. Erlbaum Associates Hillsdale, NJ, 1980), Vol. VIII, pp. 277-296.
3. Bonnel, A. M. & Miller, J. Attentional effects on concurrent psychophysical discriminations: investigations of a sample-size model. *Percept. Psychophys.* 55 (2), 162-179 (1994).
4. Wilken, P. & Ma, W. J. A detection theory account of change detection. *Journal of Vision* 4, 1120-1135 (2004).
5. Rosch, E. Cognitive reference points. *Cognit. Psychol.* 7, 532-547 (1975).
6. Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* 390, 279-281 (1997).
7. Wheeler, M. & Treisman, A. M. Binding in short-term visual memory. *J. Exp. Psychol. Gen.* 131, 48-64 (2002).
8. Jiang, Y., Olson, I. R., & Chun, M. M. Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory & Cognition* 2, 683-702 (2000).
9. Vogel, E. K. & Machizawa, M. G. Neural activity predicts individual differences in visual working memory capacity. *Nature* 428, 748-751 (2004).
10. Vogel, E. K., McCollough, A. W., & Machizawa, M. G. Neural measures reveal individual differences in controlling access to working memory. *Nature* 438, 500-503 (2005).
11. Xu, Y. & Chun, M. M. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* (2006).
12. Todd, J. J. & Marois, R. Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428, 751-754 (2004).
13. Cowan, N. *et al.* On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognit. Psychol.* 51, 42-100 (2005).
14. Zahn, C. T. & Roskies, R. Z. Fourier Descriptors for Plane Closed Curves. *IEEE Transactions on Computers* C21 (3), 269-281 (1972).

Supplementary Table 1. Adjusted r^2 values indicating the goodness of fit for the slots+averaging model when applied to the group data, the mean of the individual data, and the simulated individual data.

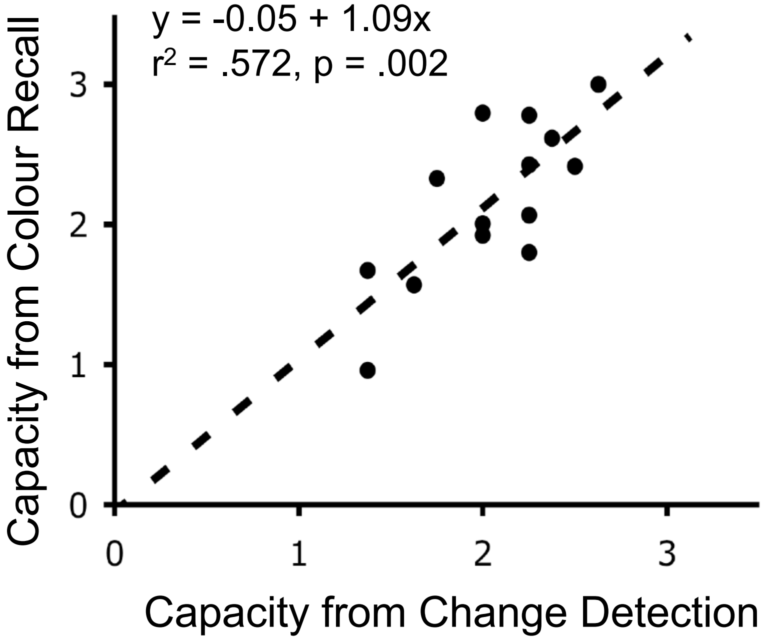
	Condition	Group r^2	Mean Individual r^2	Simulated Individual r^2
Experiment 1	Set Size 3	0.992	0.965	0.976
	Set Size 6	0.980	0.847	0.870
Experiment 2	Set Size 1	0.979	0.973	0.965
	Set Size 2	0.995	0.974	0.983
	Set Size 3	0.995	0.962	0.975
	Set Size 6	0.959	0.822	0.866
Experiment 3	Valid	0.993	0.963	0.972
	Neutral	0.986	0.835	0.906
	Invalid	0.964	0.767	0.596
Experiment 4	110-ms SOA	0.956	0.773	0.861
	340-ms SOA	0.983	0.881	0.949



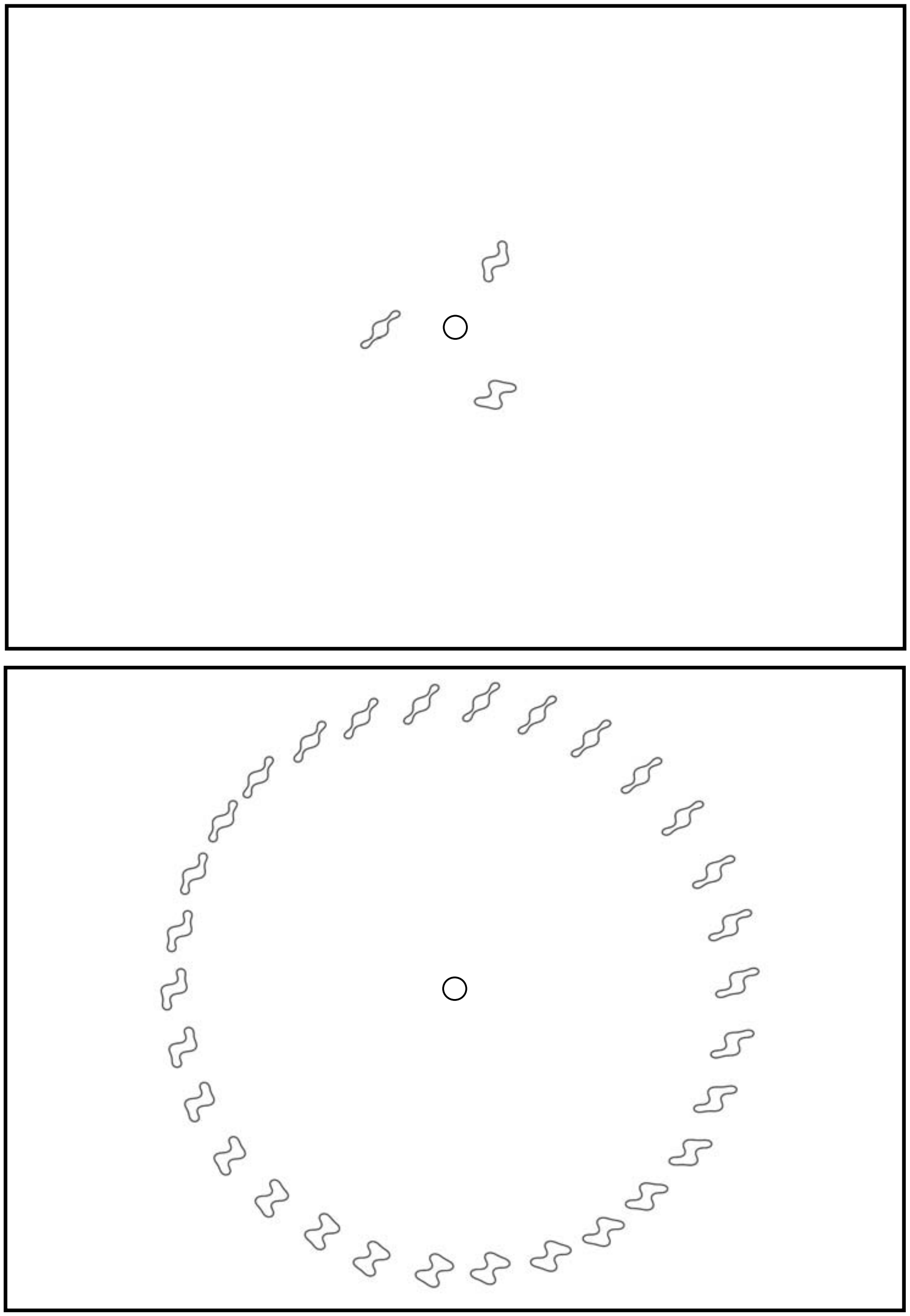
Supplementary Figure 1. Stimuli (a) and results (b) from a control experiment showing that adding sensory noise to the sample array increases the estimated standard deviation (SD) but not the estimated probability of memory (Pm). Error bars show within-subjects 95% confidence intervals.



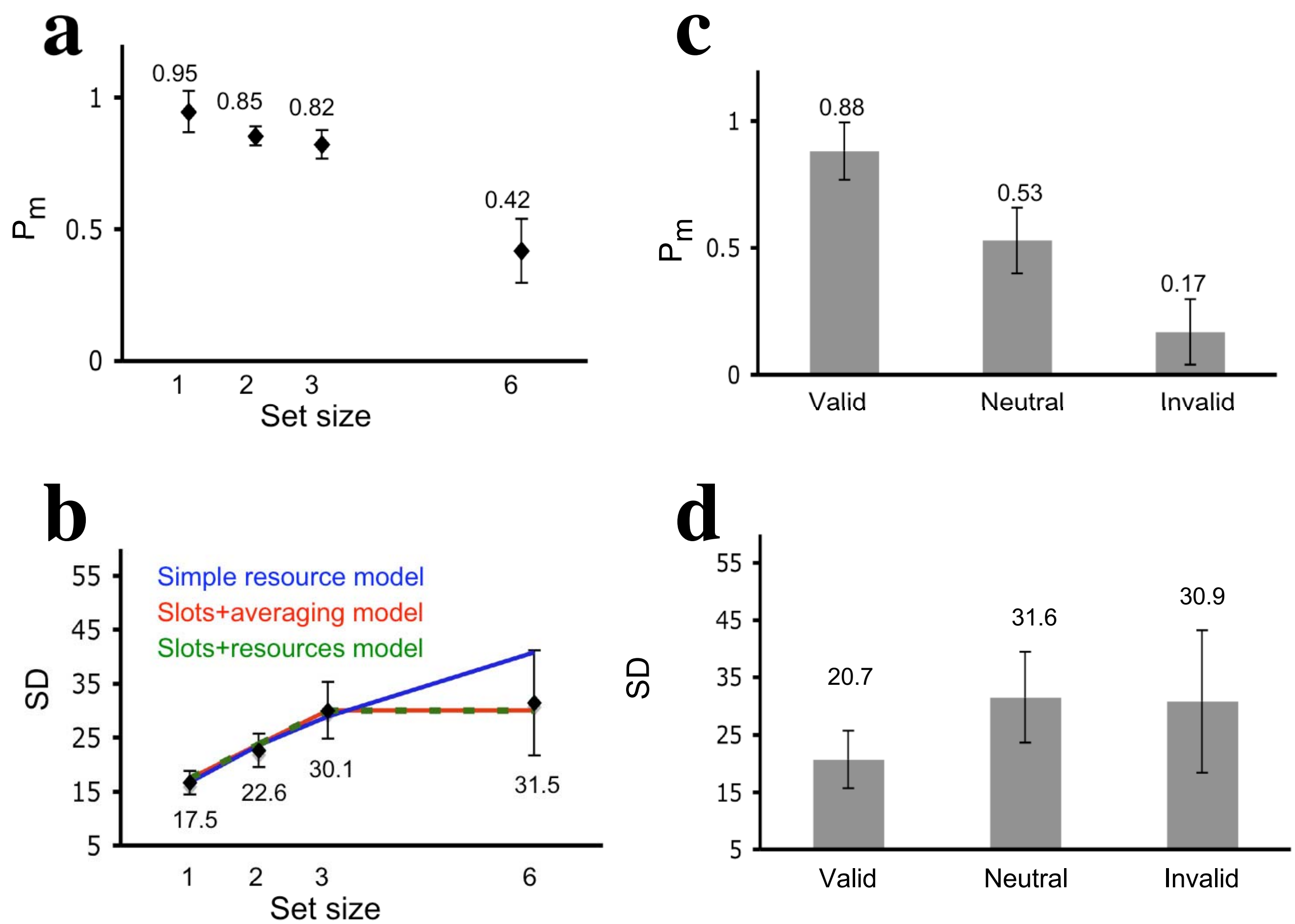
Supplementary Figure 2. **a**, Frequency of reporting a given colour value as a function of the actual colour value, aggregated across at least 150 trials for each of 50 observers across four separate experiments at set size 1. The intensity at a given point in the plot represents the frequency of occurrence for that particular combination of actual colour and reported colour. The function is nearly perfectly linear, indicating that the observers stored a representation of the actual colour (plus noise) rather than storing the nearest colour category. **b**, Analogous results from a Monte Carlo simulation of a memory system in which the actual colour value was stored as the nearest of seven equally spaced colour categories. On each trial of the simulation, Gaussian noise was added to the actual colour value, the nearest categorical value was chosen, and then the categorical colour value was reported (plus additional Gaussian noise to represent response variability). The noise levels were chosen to produce results that matched the overall level of response error exhibited by the observers in **a**. Additional simulations showed that a categorical model could not achieve the low level of response error and the linearity of the data shown by the observers in **a** unless the number of colour categories was unrealistically large (~20).



Supplementary Figure 3. Memory capacity as estimated from the colour recall task (K_i) as a function of memory capacity as estimated from a change detection task (Cowan's K).



Supplementary Figure 4. Example sample array and circle of shapes from the shape experiments. The contour of each shape can be described by the sum of two sine waves. Note that the shapes in the circle were evenly spaced in phase space, with a starting phase that varied randomly from trial to trial. The actual sample shapes were not necessarily among the exemplars shown in the circle, and the observers were instructed to make interpolated responses when the remembered shape fell between two of the exemplars in the circle. The observers were given extensive practice with making interpolated responses during a familiarization phase, in which the central sample shapes were presented simultaneously with the circle of test shapes.



Supplementary Figure 5. **(a)** P_m and **(b)** SD results from an experiment with shape stimuli in which the set size varied between values of 1, 2, 3, and 6. The lines in **b** show the predictions of the simple resource model, the slots+averaging model, and the slots+resources model. **(c)** P_m and **(d)** SD results from an experiment with shape stimuli in which a cue was presented in the sample array to indicate which item was most likely to be tested. Error bars show within-subjects 95% confidence intervals.