

The separable effects of feature precision and item load in visual short-term memory

Simon D. Lilburn

Melbourne School of Psychological Sciences,
The University of Melbourne, Parkville, Australia



Philip L. Smith

Melbourne School of Psychological Sciences,
The University of Melbourne, Parkville, Australia



David K. Sewell

Melbourne School of Psychological Sciences,
The University of Melbourne, Parkville, Australia
School of Psychology, The University of Queensland,
St Lucia, Australia



Visual short-term memory (VSTM) has been described as being limited by the number of discrete visual objects, the aggregate quantity of information across multiple visual objects, or some combination of the two. Many recent studies examining these capacity limitations have shown that increasing the number of items in VSTM increases the frequency and magnitude of errors in a participant's recall of the stimulus. This increase in response dispersion has been interpreted as a loss of precision in an item's representation as the number of items in memory increases, possibly due to a change in the tuning of the underlying representation. However, increased response dispersion can also be caused by a reduction in the total memory strength available for decision making as a consequence of a reduction in the total amount of a fixed resource representing a stimulus. We investigated the effects of load on the precision of memory representations in a fine orientation discrimination task. Accuracy was well captured by extending a simple sample-size model of VSTM, using a tuning function to account for the effect of orientation precision on performance. The best model of the data was one in which the item strength decreased progressively with memory load at all stimulus exposure durations but in which tuning bandwidth was invariant. Our results imply that memory strength and feature precision are experimentally dissociable attributes of VSTM.

is still intrinsically visual in nature, meaning that it is not simple sensory persistence but is not yet a categorically encoded representation. This posticonic, precategorical memory system is known as *visual short-term memory* (VSTM) or *visual working memory*, and has been central in the investigation of visual perception since its description by Phillips (1974).

Many of the debates within the VSTM literature have centered on different capacity constraints that may act upon this memory system and what these constraints reveal about its underlying structure. One of the central concepts in this literature is that of VSTM precision. Precision has been characterized in different ways theoretically and measured in different ways empirically, but the common assumptions that underlie research in this area are, first, that precision changes with VSTM load and, second, that loss of precision leads to poorer memory performance. Our aim in this study is to test between two alternative accounts of the psychophysical basis of precision. Both of these accounts, which were first distinguished as theoretical alternatives by Prinzmetal, Amiri, Allen, and Edwards (1998), assume that representations of stimulus attributes in memory are encoded by populations of labeled detectors or tuned channels, whose response characteristics likely reflect the characteristics of the associated neural population code. One model holds that changes in precision are associated with changes in the range or variance of the set of detectors that are activated by a stimulus feature; the other model holds that the range of detectors activated by a feature remains constant and that changes in precision are associated with changes in the strength of the active detector responses. We test between these two accounts

Introduction

Visual information can be retained for short intervals in a form that survives backward masking but

Citation: Lilburn, S. D., Smith, P. L., & Sewell, D. K. (2019). The separable effects of feature precision and item load in visual short-term memory. *Journal of Vision*, 19(1):2, 1–18, <https://doi.org/10.1167/19.1.2>.

<https://doi.org/10.1167/19.1.2>

Received February 26, 2018; published January 10, 2019

ISSN 1534-7362 Copyright 2019 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.



Downloaded From: <https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/937737/> on 01/10/2019

by investigating how the tuning of detectors is affected by changes in the number of items to retain in memory. To foreshadow our results: We find that changes in memory load lead to changes in VSTM strength, with no changes in the associated tuning function, consistent with the constant-range account. Our results imply that the loss of item precision associated with increasing memory load is better characterized as a loss in total memory strength rather than as an increase in the “blurring” or variability of item representations.

Following Luck and Vogel (1997), investigations into capacity constraints have often examined the effect of memory load, as measured by the number of visual objects presented, on overall performance. Many of these studies have employed a change detection paradigm, designed to measure VSTM quality by an observer’s ability to detect a change between a memory array and a subsequent probe array (e.g., Luck & Vogel, 1997; Phillips, 1974; Vogel, Woodman, & Luck, 2001). The item-based limitations are clear: Models with performance dependent on the total number of items have provided strong accounts of the decrease in overall performance with an increase in the number of items (Rouder et al., 2008; Sewell, Lilburn, & Smith, 2014; Zhang & Luck, 2008); constraints on retrieval time scale with the number of encoded items (Sewell et al., 2016); successively reporting features of different objects produces larger performance costs than successively reporting features of the same object (Egley, Driver, & Rafal, 1994; Woodman & Vecera, 2011); and performance seems more related to the number of items to be stored rather than the number of spatial locations (Lee & Chun, 2001; Woodman, Vecera, & Luck, 2003), mirroring similar results in the attentional literature (e.g., Duncan, 1984). Limitations have been found at the feature level, but their capacity constraints have yet to be simply characterized. Initial findings indicated that storage is contingent on feature complexity (Alvarez & Cavanagh, 2004; Eng, Chen, & Jiang, 2005) but that interitem similarity and the “resolution” of items within memory may also play a role (Awh, Barton, & Vogel, 2007; Barton, Ester, & Awh, 2009). More recent studies have also suggested that performance may be dependent on interitem, interfeature interactions (Brady & Alvarez, 2011; Brady, Konkle, & Alvarez, 2009), leading to hierarchical accounts of item–feature storage (Brady et al., 2011; Orhan & Jacobs, 2013).

These later studies, along with many others, employed a continuous report paradigm (e.g., Bays, Catalao, & Husain, 2009; Wilken & Ma, 2004; Zhang & Luck, 2008), which assesses an observer’s ability to reproduce the remembered value of a continuously distributed stimulus attribute like color or orientation. In its use of a continuous response variable, the continuous report paradigm is reminiscent of the

method of adjustment used in measuring thresholds in classical psychophysics (Woodworth & Schlosberg, 1954). The method was resurrected by Prinzmetal et al. (1998) as a way to assess the effects of attention on perceptual variability and, following the work of Wilken and Ma (2004), has been widely used in VSTM studies to assess representational precision in memory. Studies employing a continuous report paradigm indicate that although an item limit might be a critical constraint on the overall memory system, there is a systematic increase in the observer error rate below any item limit, which can be seen when tasks require that small changes in a feature be detected (Bays et al., 2009; Palmer, 1990; Wilken & Ma, 2004) or a feature be reproduced from a memory representation (Wilken & Ma, 2004; Zhang & Luck, 2008). In this paradigm, errors in which the observer reproduces a similar, but not identical, stimulus are generally modeled as a Gaussian or von Mises distribution (a circular approximation of the Gaussian distribution) centered on the true stimulus identity. A consistent finding is that even below the item limits identified in the change detection literature, the variance of the distributions of response errors increases as a function of the memory array size. This indicating that observers are more likely to produce stimuli increasingly different from the target stimulus with additional items (or to mistake stimuli increasingly different from the target as having been previously presented). Whether a hard item limit can be reconciled with this finding has been a continuing point of debate (Adam, Vogel, & Awh, 2017; van den Berg & Ma, 2014; Zhang & Luck, 2008).

Taken together, these change detection and continuous report data indicate that VSTM performance is highly contingent on the number of items presented to the observer, but that a simple item limit does not appear to be the complete picture. One difficulty is that the change of performance observed, even at small item loads, across feature values within continuous report tasks does not easily map onto a single type of memory constraint. Although precision is often invoked as a singular concept in discussions of continuous report tasks, it refers to both an observable property of responding (the inverse of response error variance) and a property of the resolution of the underlying representation across values of a stimulus feature (sometimes called *mnemonic precision*; see, e.g., van den Berg, Awh, & Ma, 2014). Despite being obviously existent, the exact relationship between these two quantities is not necessarily straightforward, as evidenced by the large number of candidate models that are broadly consistent with continuous report data (Bays et al., 2009; Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011; van den Berg et al., 2014; van den Berg, Shin, Chou, George, & Ma, 2012; Zhang & Luck, 2008). An increase in response dispersion (that is, a

decrease in response precision) could be consistent with a constant change in total memory strength—the signal-to-noise ratio *across* all values of a feature—as argued for by Palmer (1990), Smith and Sewell (2013), Bays (2014), and Smith (2016). There are multiple possible causes for such a change, including dynamic remapping of an existing resource (Bundesen, Habe-kost, & Kyllingsbæk, 2011; Ester, Serences, & Awh, 2009) and an increase in interference between items simultaneously stored within memory (Oberauer & Lin, 2017).

Another possibility is that changes in the number of items to be retained also change the representation *within* the domain of feature values. Change in the representation-level precision of an item as a function of the total number of items could be the result of averaging across multiple noisy copies of an item stored in memory (Zhang & Luck, 2008) or a change at the fundamental level of detector responses, so-called *nonmultiplicative scaling* (Reynolds & Heeger, 2009), conjectured by Prinzmetal et al. (1998) and observed at the neural level in work pertaining to the underlying mechanisms of attention (Martinez-Trujillo & Treue, 2004; Spitzer, Desimone, & Moran, 1988; Treue & Martinez-Trujillo, 1999). We are not committed to any specific mechanism for changes within or across the feature domain but are more generally interested in the interaction between item-level constraints on performance and feature-level constraints at the level of psychophysical channels or detectors (Graham, 1985).

Evidence that can distinguish between different sets of limitations can be obtained by placing additional experimental constraints on observers and imposing more specific constraints on modeling observer performance. In the current article, we use a simple two-choice discrimination task with well-known decision properties to examine observer memory for stimuli with small angular differences (target angular offsets) from a known referent (in this case, a vertically oriented stimulus). Our previous work examining memory performance using this paradigm has provided a strong theoretical constraint, an inverse square-root relationship between memory load and detection sensitivity d' (Palmer, 1990), which can be used to provide a quantitative characterization of the relationship between item- and feature-level capacity constraints.

A sample-size information constraint

Changes across memory array sizes in the variance of reproduction error within continuous report tasks and in the precision with which observers can distinguish small changes have been described as following a power law, where precision (defined as either the inverse variance or the inverse standard deviation of the error

distribution) for a display of m items changes with the relationship $P_m = P_1 m^{-k}$, where the subscript of the precision P denotes the memory array size (Bays & Husain, 2008; Donkin, Kary, Tahir, & Taylor, 2016; Palmer, 1990; Smith, Corbett, Lilburn, & Kyllingsbæk, 2018). More concretely, this power law can be interpreted as the division of a fixed memory resource across all items to be remembered. The exponent k of this power law appears to reflect whether this division of attentional resources across items in the display is balanced or uneven (Smith et al., 2018).

In the simplest cases known—discrimination of length between line segments (Palmer, 1990) and discrimination between Gabor patches that are orthogonally oriented (Sewell et al., 2014)—an even division of resources across all items in a display leads to the relationship

$$d'_m = \frac{d'_1}{\sqrt{m}},$$

where d' is the sensitivity measure from signal detection theory, m is the display size, and the subscript again denotes the corresponding memory array size. This capacity constraint, characterized by a sample-size model of VSTM resource allocation, was identified in the orthogonal orientation discrimination case in our earlier work (Sewell et al., 2014, 2016). Although this capacity limitation has been identified as a limitation in consolidation by Vogel et al. (2006), we reported that the same limitation could be obtained regardless of whether individual items were presented simultaneously or sequentially. We also found that a model which assumed that this performance constraint was due to an informational capacity limit of the memory system accounted for data better than an object-level exponential race model which assumed that the capacity limit was due to a limitation in the rate of consolidation (Sewell et al., 2014).

Such an information limitation can be reformulated as the result of aggregating a fixed-size pool of opponent-coded Poisson neurons (Smith, 2015) or, more generally, as the normalization of the strength of information within memory based upon stimulus energy (Smith & Sewell, 2013; Smith, Sewell, & Lilburn, 2015). This constraint makes clear predictions for performance across different memory array size conditions, as additional items place additional demands (potentially equal demands) on a fixed information limit. The sample-size relationship does not, however, place any constraints on how that information might be distributed across feature values (in this case, different orientation values). An account of fine orientation discrimination performance must generalize this information constraint to also include any effect that discriminability of the feature values may have on overall performance.

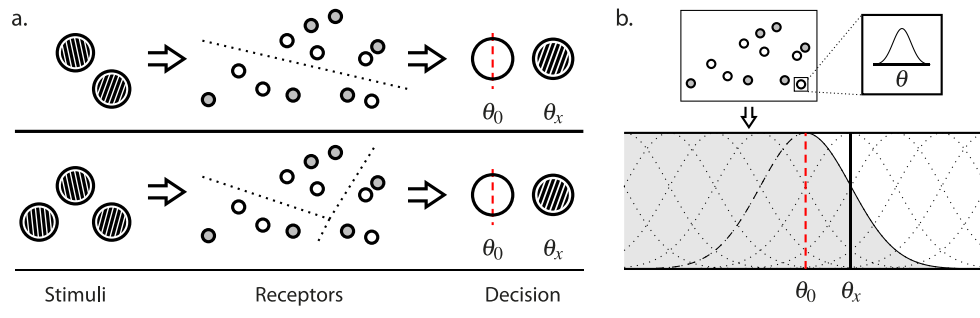


Figure 1. A schematic overview of the sample-size/tuned channel mechanism proposed in accounting for the effect of object-level and feature-level constraints on observer performance in a fine orientation discrimination task. (a) Three stages in the fine orientation discrimination task: encoding the stimulus; representing the maintained stimuli by a fixed pool of receptors; and deciding about the orientation of a probed representation (θ_x) to a known (vertical) stimulus standard (θ_0). The sample-size/tuned channel model posits that stimuli are represented as a fixed pool of orientation-specific receptors, some corresponding to orientations clockwise from a vertical orientation (dots filled with white) and some counterclockwise (dots filled with light gray). These receptors are divided between different stimuli. (b) The discriminability of an orientation decision from a population of detectors. Each detector has some response to a range of orientations, described by its tuning function. Assuming dense and uniform coverage of all orientation values by detectors, the sensitivity of a decision is determined by the ratio of the response of all detectors with correct response labels (white, in this case) to the response of detectors with incorrect response labels (light gray) for a given orientation (θ_x).

Tuned channels within memory representations

It is assumed, in discriminating between two orthogonal orientations, that the two orientations—horizontal and vertical, in the case of Sewell et al. (2014)—are maximally discriminable. This is in keeping with psychophysical investigations of discrimination performance involving successively presented near-threshold gratings, where maximum performance is reached when gratings differ by more than 20°–30° (Thomas & Gille, 1979). Around and below this point, discrimination performance in a memory task would be limited by the information capacity of VSTM and by sensory encoding constraints, as well as any interaction between the two (caused by, for instance, changes in the precision of a representation within memory).

To model the effect of target angular offset on observer performance, we assume that orientation is stored within memory as activity in a set of overlapping tuned channels, or labeled detectors (Graham, 1985; Watson & Robson, 1981), each of which responds in a manner proportional to the agreement of the current memory representation and some internal orientation (see Figure 1). In the current two-alternative case, where observers must only indicate whether the target was rotated clockwise or counterclockwise relative to a vertical orientation, it is only necessary to consider two sets of channels which must be integrated for any single decision: those that code for the correct response (the matching detectors) and those that code for the incorrect response (the nonmatching detectors). Only the signal-to-noise ratio of the response for correct and incorrect detectors matters, and the critical region for examining changes in this ratio is where the responses

from the two sets overlap: where the detectors associated with the nonmatching detectors also respond to stimuli with orientations associated with the matching response. The bandwidth of the tuning curve for detectors immediately on either side of the referent orientation then determines the rate at which performance will change as target angular offset increases. Lower bandwidth produces sharper tuning curves, resulting in more rapid performance gains with increasing angular offset.

The most straightforward way to combine the constraints imposed by the sample-size model with those imposed by a tuned channel model is to multiplicatively combine them. This way of combining the constraints makes the assumption that the effects of VSTM load on channel tuning (the underlying variance in the representation of an item) and the strength of the item representation are separable, with the changes in the number of items affecting either or both of the different constraints. In order to combine the sample-size relation with a tuned channel model, we assume that there exists some latent maximum level of performance that would be seen in the case of an observer making a discrimination of a single item in memory that was maximally discriminable from the referent for a display exposed for a single unit of time (e.g., 1 s). This maximum level, which we denote as $d'_{1,1,\pm\pi/4}$, functions as a scaling parameter of the model in the sense that the effects of display size, target angular offset, and stimulus exposure duration are all expressed as functions of this d' . The first and second subscripts denote the display size and stimulus exposure duration, while the third indicates that the

maximum d' should be achieved when stimuli are oriented at $\pm\pi/4$ radians ($\pm 45^\circ$) to the vertical.

For consistency with the properties of the sample-size model, we assume that the function that characterizes the effect of angular difference on performance weights the *squared sensitivity* of the maximally discriminable stimulus difference rather than sensitivity itself. According to the sample-size model, performance is predicted to be inversely proportional to the square root of the number of items to be retained, a consequence of basic sampling theory (the standard error of a sample mean is inversely proportional to the square root of the number of samples taken, which are then evenly divided among each of the items in memory). An extension of the sample-size model to the simplest separable tuned channel model would likewise characterize the effects of channel tuning at the level of squared sensitivity. Direct changes in squared sensitivity are also seen in the model of human performance in visual tasks involving multiple near-threshold items developed by Smith and Sewell (2013). In that model, memory performance is determined through the normalization of stimulus energy rather than the normalization of stimulus strength or amplitude (the latter being the square root of the former). In that original presentation of the model, the stimulus-identity information is carried in the sign of the process, with a value of zero interpreted as a memory trace without discriminating information for the current decision. This reflected the idealization that orthogonally oriented stimuli carry equal and opposite quantities of information to one another. Our model presented in this article extends that previous model to include nonorthogonally oriented stimuli by weighting squared sensitivity inversely to the total number of total items and proportionally to the discriminability of orientation information.

Predicted sensitivity in the model depends on the distance between the target (the probed stimulus) and an internal referent assumed to be on the boundary between detectors associated with matching and nonmatching responses: Performance should be at chance when the target and referent are identical and will increase with distance between the two until maximum performance is reached. In our experiment, in which observers made judgments about whether stimuli were oriented clockwise or counterclockwise from vertical, we assume the internal referent is a completely vertical stimulus (what would be a stimulus of zero discriminability, were it to be presented). The distance that determines performance is the angular difference between the target and verticality transformed through a tuning function representing the activity of a nonmatching detector centered on a vertical orientation.

The amount of discriminable information is determined by the distance of the stimulus feature away from the tuning envelope of the nonmatching detector, with the discriminability of a stimulus decreasing as the response of the nonmatching detector increases, leading to the multiplicative term $1 - f[b(\theta - \theta_0)]$. Here $f[\cdot]$ is the tuning function of the channel with an assumed range of 0.0 to 1.0, b is the bandwidth of the channel, θ is the angle of the target stimulus, and θ_0 is the angle that the tuning function is centered on. We use a conventional Gaussian-shaped tuning curve (Vogels, 1990; Zemel, Dayan, & Pouget, 1998), where $f(x) = \exp(-x^2)$, although we have found that a cosine or other smooth-shaped function can be substituted with no change in quality of the fit to empirical data (in line with the observation of Thomas & Gille, 1979).

A key question for the current experiment is whether the bandwidth of the tuning function is independent of the display size or varies with the number of items in the display, indicating an interaction between feature-level precision within memory and the number of items simultaneously maintained. We investigated this in our modeling by comparing models in which the bandwidth parameter b was the same for all display sizes with models in which it varied as a function of display size.

By combining simple tuned channel weighting with the sample-size model we can predict the performance for each experimental condition, denoted by the triple $\langle m, \tau, \theta \rangle$ (memory array size, stimulus exposure time, and target angular offset), as

$$d'_{m,\tau,\theta} = \frac{\sqrt{1 - f[b(\theta - \theta_0)]} d'_{1,1,\pm\pi/4} \tau}{\sqrt{m}}$$

The third constraint on observer performance is the exposure duration of the memory array, represented as a simple multiplicative term in the equation. A summary of the possible free parameters is provided in Table 1.

In the model fits presented in the following, we fit both models in which time weights sensitivity (following the equation just given) and models in which it weights squared sensitivity (replacing τ with $\sqrt{\tau}$ in the equation). The latter form is expected through a strong mechanistic interpretation of the sample-size model, which assumes that samples are recruited at a constant rate (Sewell et al., 2014), but the former seems to provide a better account of the current data. Although a strong interpretation of this term would suggest that observer performance would increase indefinitely with longer stimulus exposure durations, we would expect that such growth would slow down or stop as the limited pool of resources—implied by the sample-size constraint—is exhausted. This pattern of reaching asymptotic performance with longer stimulus exposure durations was demonstrated by Bays et al. (2011), who found that observer performance leveled off beyond

Parameter	Description
$d'_{1,1,\pm\pi/4}$ *	<i>Latent maximal sensitivity.</i> When the sample-size constraint is enforced, only a single sensitivity parameter is used across memory array sizes.
$d'_{1,1,\pm\pi/4}, \dots, d'_{m,1,\pm\pi/4}$	<i>Sensitivity for memory array size m.</i> When the sample-size constraint is not enforced, a sensitivity parameter is estimated separately for each memory array size.
b^*	<i>Single bandwidth.</i> A single bandwidth shared across all conditions.
b_0	<i>Bandwidth intercept.</i> The intercept component of a linear relationship between memory array size and target angular offset performance.
b_i	<i>Bandwidth slope.</i> The change in the bandwidth of the tuning function with a change in memory array size.
b_1, \dots, b_m	<i>Bandwidth for memory array size m.</i> A freely estimated bandwidth for display size m .
Bias	Response bias was fixed at zero or a parameter was used to account for response bias (one parameter shared across all response categories or a parameter estimated for each response category).

Table 1. Free parameters used in model construction. *Notes:* Parameters marked with an asterisk (*) were in the two-parameter model that was ranked first in terms of Bayesian information criterion for the group-average data. For the group-average data, no response bias was required. For individual observers, a single bias parameter for all response categories was also required.

200–300 ms of stimulus exposure. Given that there is a greater risk of verbal recoding and categorization of stimuli with longer exposure durations and retention intervals, we have chosen to examine a limited regime of stimulus exposure durations where an information constraint is still apparent. Further comments regarding this term will be saved for the Discussion.

Each of the three constraints—the sample-size constraint, the tuned channel constraint, and the exposure-time constraint—can enter the model as completely independent terms without interactions between the three corresponding experimental manipulations (respectively, the number of items, the angular offset of the target stimulus, and the stimulus exposure duration).

Method

Participants

Five observers (HA, LA, ME, SN, and TS) participated in this study, all paid observers selected from the University of Melbourne who were unaware of the aims of the study. Each observer was briefed about the general nature of the study, signed a consent form prior to participation, and was remunerated AUD \$12 for each session completed. Each observer completed a variable number of practice and calibration sessions to gain familiarity with the stimuli and control task difficulty through the per-observation manipulation of stimulus contrast. Observers completed 18 to 20 sessions in total, of which 15 were experimental sessions. Each session lasted approximately 35 min, with regular breaks between blocks of trials for observers to rest.

Stimuli and apparatus

The stimuli used in the memory and probe arrays were oriented Gabor patches: Gaussian vignettted 3.5 c/° sinusoidal luminance gratings subtending 0.97° of visual angle at half height. The form of the Gabor patches was as given by Graham (1989, p. 53). Both the target and distractor patches were oriented (with equal probability) either clockwise or counterclockwise by 0.1, 0.3, 0.5, or 0.7 radians (i.e., 6°, 17°, 29°, or 40°) away from a vertical position and placed on a field of mean luminance 30 cd/m². Patches could be presented at four locations, diagonally located at a distance of 2.3° from a central fixation cross subtending 0.29° of visual angle. To maximize the effect of stimulus duration on memory formation, the period required to encode the stimuli was extended by embedding the stimuli in dynamic noise (Lu & Doshier, 1998; Ratcliff & Smith, 2010; Sewell et al., 2014; Smith, Ratcliff, & Sewell, 2014). The noise patches were composed of blocks of luminance 4 × 4 pixels in size sampled from a truncated Gaussian distribution with a mean of the background luminance and a variance scaled to fit within 20% of the entire luminance range of the display. Noise patches were displayed on alternating frames to the stimulus display during the stimulus-display period, meaning that each 10 ms of stimulus display was immediately followed by 10 ms of noise. A high-contrast bull's-eye backward mask was used to terminate stimulus presentation and disrupt any sensory memory trace. Stimuli were generated on a Cambridge Research System ViSaGe frame store and presented on a gamma-corrected 21-in. Sony Trinitron Multiscan G520 monitor, running at a resolution of 1,024 × 768 pixels and driven at 100 Hz (giving a frame duration of 10 ms). Custom C++ software was used to generate the stimuli, control trial presentation, and

record responses. Observers performed the task in a dimly lit observation booth at a viewing distance of 100 cm. Viewing position was stabilized with a chin rest.

Procedure

A $4 \times 3 \times 4$ within-subject design was used, composed of four memory array sizes (1–4 items), three stimulus exposure durations (100, 150, and 200 ms), and four target angular offsets. Each session of the experiment consisted of 384 trials, yielding a total of 5,760 trials per observer after the full 15 experimental sessions. All stimuli were presented at a single level of contrast for each observer. This contrast level was selected individually for each observer during the practice sessions to provide the maximum range between the most difficult condition (four items presented for 100 ms, with the target very close to a vertical orientation) above chance, and the least difficult condition (a single item presented for 200 ms, with the target item oriented 0.7 radians away from a vertical position) below ceiling performance.

Each trial consisted of the presentation of a uniform gray field for the 1,000-ms foreperiod, followed by the presentation of a fixation cross for 1,500 ms. The memory array was then presented, containing one to four items, for either 100, 150, or 200 ms. The presentation of the memory array was terminated with the presentation of high-contrast bull's-eye masks for 200 ms in locations corresponding to the memory array stimuli. After this 200 ms backward masking period, a report cue was displayed at one of the memory array locations, indicating that the observer was to judge whether the memory stimulus in this location was clockwise or counterclockwise from vertical. This report display was shown until a response was entered. Audible feedback was then presented to the observer to indicate the accuracy of the response. A schematic of the stimuli and the presentation regime can be seen in Figure 2.

Results

The proportions of correct and error responses were aggregated for each response category (responding that the target was clockwise or counterclockwise of a vertical orientation) and each experimental condition (memory array size, exposure duration, and target offset) for each observer. A group average was also computed for each response category and experimental condition.

A set of 108 different models was constructed and fitted to each observer and the group average, with the

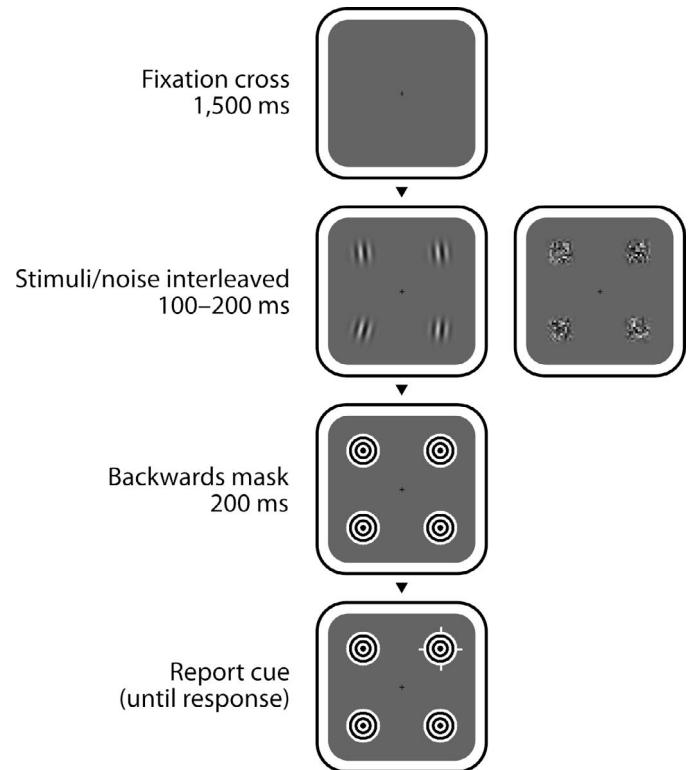


Figure 2. A schematic of a single trial of the experiment: Stimuli are presented, interleaved with patches of Gaussian-distributed noised, and followed by the presentation of high-contrast bull's-eye masks and then the report cue (marker around the top right stimulus).

models differing in five ways. First, the relationship between sensitivity and memory array size was either freely estimated or constrained to the sample-size relationship (denoted “SS constraint” in the fit tables). Second, the relationship between memory array size and the bandwidth of the tuning function was either fixed to a single estimated value, freely estimated for each memory array size, or constrained to be a linear function of memory array size (denoted “*b* interaction” in the fit tables). Third, the center of the tuning function was either fixed to the point of zero discriminability (i.e., a completely vertical stimulus) or estimated freely (denoted “Offset” in the fit tables) to allow for nonlinearities in the psychophysical response around the response boundary. Fourth, sensitivity was constrained to increase either linearly or quadratically as a function of stimulus exposure duration, or the effect of stimulus exposure duration on sensitivity was freely estimated (denoted “Growth” in the fit tables). And fifth, the response bias (for responding either clockwise or counterclockwise) was fixed to unbiased optimal responding, estimated as a single response-bias parameter for all experimental conditions, or estimated

Rank	SS constraint	<i>b</i> interaction	Offset	Growth	Bias	<i>k</i>	G^2	BIC
1	Yes	Zero	Zero	Linear	0	2	37.210	54.528
2	Yes	Linear	Zero	Linear	0	3	34.487	60.463
3	Yes	Zero	Zero	Linear	1	3	37.008	62.984
4	Yes	Zero	Free	Linear	0	3	37.086	63.062
5	Yes	Linear	Free	Linear	0	4	34.150	68.785

Table 2. Goodness-of-fit statistics for the top five best-fitting models for group-average data. *Notes:* SS constraint = whether the sample-size constraint was imposed; *b* interaction = interaction between tuning function bandwidth and memory array size; Offset = the center of the tuning function; Growth = the growth of sensitivity over time; Bias = whether response bias was freely estimated (either for every condition or a single bias for the observer); *k* = the number of estimated parameters; BIC = Bayesian information criterion.

separately for each response category (denoted “Bias” in the fit tables).

For each model, a G^2 goodness-of-fit statistic was computed:

$$G^2 = 2 \sum_i N_i \sum_j p_{ij} \ln \left(\frac{p_{ij}}{\pi_{ij}} \right),$$

where *i* indexes the experimental condition (across all conditions of target angular offset, memory array size, and stimulus exposure duration), *j* indexes the response outcome (i.e., correct or incorrect responses), N_i is the number of observations in the *i*th experimental condition, p_{ij} is the observed proportion of *j* responses in the *i*th experimental condition, and π_{ij} is the predicted proportion of *j* responses in the *i*th experimental condition. Each model configuration was fitted to each observer and the group-average data 10 times using the Nelder–Mead simplex method, initiated with 10 different starting locations. The Bayesian information criterion (BIC) was computed to compare the goodness of fit while also taking into account model complexity: $\text{BIC} = G^2 + k \ln(N)$, where *k* denotes the number of estimated parameters and *N* denotes the total number of observations.¹ We have chosen to use the BIC for model selection because it emphasizes model parsimony. Other measures, such as the Akaike information criterion, are known to increasingly favor more complex models as sample size increases. Consistent with this, the Akaike information criterion—when applied to our data—preferred more complex models for all observers, but there was no systematic or interpretable pattern to the differences between the model orderings with the different information criteria. We have chosen the BIC because of the theoretical simplicity and parsimony of the picture it provides.

The group-average data were treated statistically as if they were data from a single typical observer. Although model fits to group data are not based on true likelihoods, we have nevertheless found such fits useful because they emphasize the common features of performance across individuals to provide clearer

theoretical interpretations and deemphasize the idiosyncrasies of individual performance. In the group-average model fits, the *N* in the computation of the BIC is the average number of trials across observers (5,760). An additional joint model fit was determined by computing a G^2 for each model across all observers, and computing the BIC by summing the number of observations and parameters for all observers. For the joint model fit, the *N* in the computation of the BIC is the total number of trials across all observers ($5 \times 5,760$).

The five best-fitting models (ranked in terms of BIC) for the group-average data are presented in Table 2. The top-ranking model imposes a sample-size constraint, predicts a linear increase in sensitivity with stimulus exposure duration, and imposes a Gaussian-shaped tuning function which is invariant across conditions of memory array size (that is, there is no bandwidth interaction). The predictions of this best-fitting model for the group-average data are presented in Figure 3.

For four of the five individual observers, the same model was ranked highest in terms of BIC out of the 108 configurations tested. For the remaining observer (ME), a linear interaction between memory array size and tuning function bandwidth was found; the sample-size-constrained model without an interaction between bandwidth and memory array size was ranked 11th ($\Delta\text{BIC} = 28.002$). The five best-fitting models for each observer are shown in Table 3, and the individual data (and the best model fit for each observer) are displayed in Figure 4. Rankings for the joint model fits are given in Table 4, with the sample size constrained model without an interaction ranking top, with a slight advantage over the model with a linear interaction between tuning function bandwidth and memory array size. Although the models with and without a dependency between tuning function and memory array size rank closely together, the consistency of the BIC advantage for the more parsimonious model in four out of the five observers (as well as the group

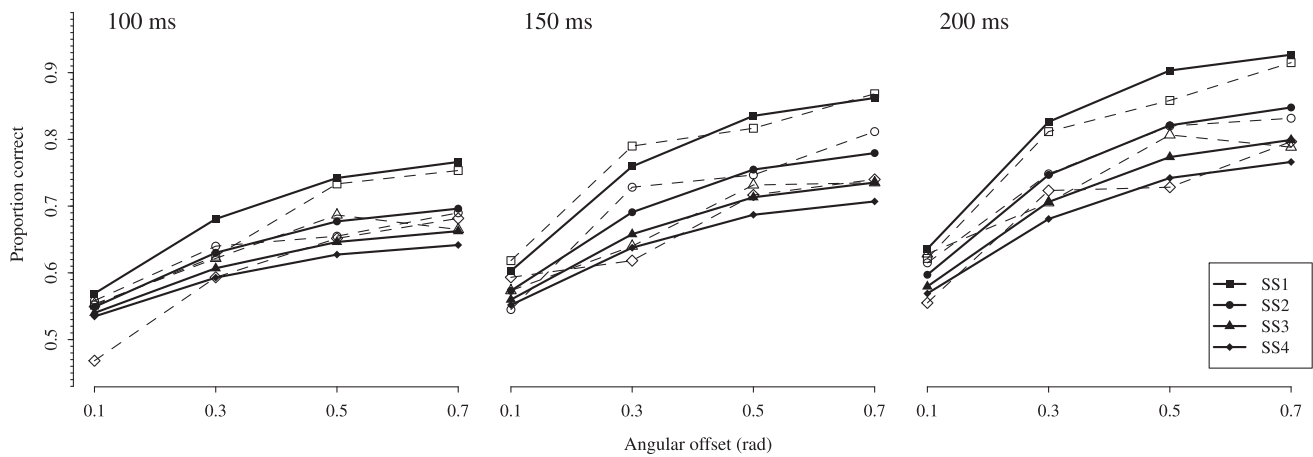


Figure 3. The preferred Bayesian information criterion model for the group-average data (displayed in solid lines with filled markers) against the group-average data (displayed in dashed lines with unfilled markers). The best-fitting model imposes the sample-size constraint, a linear increase in sensitivity with time, and a Gaussian-shaped tuning function which is invariant across display size manipulations. Each panel represents a different stimulus exposure duration (shown at the top left of the panel). SS = memory array (set) size.

Obs.	Rank	SS constraint	<i>b</i> interaction	Offset	Growth	Bias	<i>k</i>	<i>G</i> ²	BIC
HA	1	Yes	Zero	Zero	Linear	0	2	195.484	212.801
HA	2	Yes	Zero	Free	Linear	0	3	194.306	220.282
HA	3	Yes	Zero	Zero	Linear	1	3	195.261	221.237
HA	4	Yes	Linear	Zero	Linear	0	3	195.436	221.412
HA	5	Yes	Linear	Free	Linear	0	4	192.424	227.059
LA	1	Yes	Zero	Zero	Linear	1	3	161.570	187.546
LA	2	Yes	Linear	Zero	Linear	1	4	157.663	192.297
LA	3	Yes	Zero	Free	Linear	1	4	161.410	196.045
LA	4	Yes	Linear	Free	Linear	1	5	157.516	200.809
LA	5	Yes	Zero	Zero	Free	1	6	151.116	203.068
ME	1	Yes	Linear	Zero	Linear	1	4	179.974	214.609
ME	2	Yes	Linear	Free	Linear	1	5	172.250	215.543
ME	3	Yes	Free	Zero	Linear	1	6	177.558	229.510
ME	4	No	Zero	Zero	Linear	1	6	177.950	229.902
ME	5	Yes	Free	Free	Linear	1	7	169.883	230.494
SN	1	Yes	Zero	Zero	Linear	1	3	124.255	150.231
SN	2	Yes	Linear	Zero	Linear	1	4	117.412	152.046
SN	3	Yes	Linear	Free	Linear	1	5	110.013	153.307
SN	4	Yes	Zero	Free	Linear	1	4	119.695	154.330
SN	5	Yes	Zero	Zero	Squared	1	3	135.326	161.302
TS	1	Yes	Zero	Zero	Linear	1	3	195.545	221.521
TS	2	Yes	Linear	Zero	Linear	1	4	192.600	227.235
TS	3	Yes	Zero	Free	Linear	1	4	193.544	228.179
TS	4	Yes	Linear	Free	Linear	1	5	191.561	234.854
TS	5	Yes	Free	Zero	Squared	1	6	185.159	237.111

Table 3. Goodness-of-fit statistics for the top five best-fitting models for each observer. Notes: Obs. = observer; SS constraint = whether the sample-size constraint was imposed; *b* interaction = interaction between tuning function bandwidth and memory array size; Offset = the center of the tuning function; Growth = the growth of sensitivity over time; Bias = whether response bias was freely estimated (either for every condition or a single bias for the observer); *k* = the number of estimated parameters; BIC = Bayesian information criterion.

Rank	SS constraint	<i>b</i> interaction	Offset	Growth	Bias	<i>k</i>	G^2	BIC
1	Yes	Zero	Zero	Linear	1	15	893.264	1047.286
2	Yes	Linear	Zero	Linear	1	20	842.864	1048.227
3	Yes	Linear	Free	Linear	1	25	823.561	1080.265
4	Yes	Zero	Free	Linear	1	20	877.620	1082.983
5	No	Zero	Zero	Linear	1	30	815.136	1123.180

Table 4. Goodness-of-fit statistics for the top five best-fitting models for the joint data. *Notes:* SS constraint = whether the sample-size constraint was imposed; *b* interaction = interaction between tuning function bandwidth and memory array size; Offset = the center of the tuning function; Growth = the growth of sensitivity over time; Bias = whether response bias was freely estimated (either for every condition or a single bias for the observer); *k* = the number of estimated parameters; BIC = Bayesian information criterion.

average and the joint model fit) is, we believe, compelling.

Discussion

Theoretical accounts of VSTM capacity have differed in terms of which level of representation is thought to be constrained: whether the level of whole objects (as, for instance, in a “slots” model; e.g., Luck & Vogel, 1997), the level of overall information content (in a “resource” model; e.g., Bays et al., 2009; Wilken & Ma, 2004), or some combination of the two (e.g., van den Berg et al., 2014; Zhang & Luck, 2008). Our results show that, when ranked in terms of BIC, a parsimonious model combining the sample-size constraint previously reported with a tuning function fitted both group-average and individual-observer data well. The sample-size relation describes a change in performance with the number of items to be remembered, consistent with the distribution of a limited pool of neural resources between item representations. The tuning function implies that the accuracy of fine orientation discrimination decisions is limited by the specificity of orientation-specific neural detectors to oriented input. The effect of stimulus exposure duration was modeled as a simple linear increase in sensitivity with exposure time, the implications of which will be discussed in the following. These elements provide separate constraints on performance within the model.

In our current modeling, the tuning function was not dependent on the number of items held in memory: in most of the best fitting models, the overall performance of an observer did not require a change in the bandwidth of a tuning function to capture performance across different conditions of memory array size. An increase in the number of items to be retained shifted the sensitivity curve down by an inverse square-root factor predicted by the sample-size model but did not change the shape of the tuning curve. This means that, under the assumptions of this model, observer performance may be factorized into the independent effects of

memory array size, stimulus exposure duration, and feature-level discriminability.

One potential objection that could be raised to our interpretation of our results is that we have not shown that the memory representations retrieved by the report cue are truly visual in form. An alternative interpretation is that the orientations of the stimuli are rapidly classified as clockwise or counterclockwise during the stimulus exposure period, prior to the mask, and are represented in this binary, categorical form during the postmask retrieval period, when they are then cued for recall. These alternatives cannot be distinguished on the basis of response accuracy alone, because accuracy does not identify when a decision about stimulus identity is made. However, the fast-decision hypothesis can be ruled out by studies of the time course of perceptual decisions about brief masked stimuli using the diffusion decision model (Ratcliff & Rouder, 2000; Smith, Ellis, Sewell, & Wolfgang, 2010; Smith, Ratcliff, & Wolfgang, 2004). The consistent picture provided by these studies is that the time taken to make a decision about a near-threshold masked stimulus can be an order of magnitude (i.e., a factor of 10) longer than the time for which the stimulus is physically present. The response-time distributions and choice probabilities from these kinds of tasks can be well described mathematically by a model in which a stable representation of a transient stimulus event is maintained in VSTM after stimulus offset until the decision process is complete (Smith & Ratcliff, 2009). These kinds of decisions typically take from 500 to 1,500 ms or longer to complete, which is inconsistent with the hypothesis that the entire contents of the display can be classified and recoded in binary form prior to the mask.

One study that is particularly relevant to the fast-decision hypothesis is that of Sewell et al. (2016), who investigated VSTM for orthogonally oriented grating patches in an experimental paradigm very similar to the one we used here. Those researchers found that mean response times in this task varied from around 700 to 1,300 ms, depending on the stimulus exposure and the number of items in the display, with the shortest mean response times for single-item displays at the longest

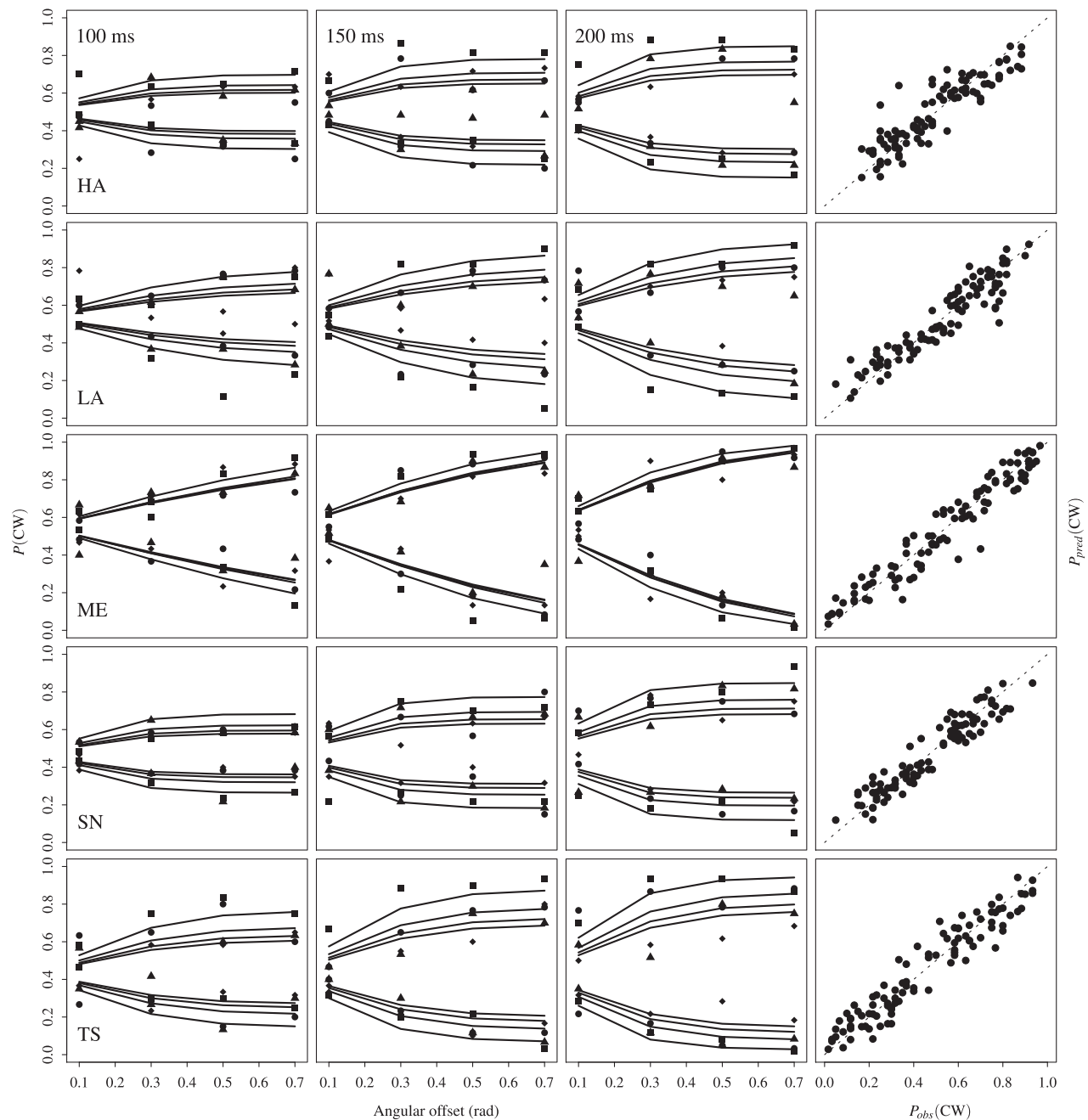


Figure 4. The preferred Bayesian-information-criterion model for each observer (displayed in solid lines) plotted against their respective data (displayed as markers according to display size). As per-observer response bias may be more severe than any group-average response bias, we display the data by the probability $P(\text{CW})$ of responding “clockwise,” rather than by aggregated proportion correct. Model predictions and observed data above $P(\text{CW}) = 0.5$ represent hits; those falling below $P(\text{CW}) = 0.5$ are false alarms (i.e., responding “clockwise” to a counterclockwise stimulus). For most observers (other than observer ME), the preferred model had a sample-size relationship between display size and performance, a linear increase in sensitivity with stimulus exposure duration, and a Gaussian-shaped tuning function which is invariant across display size manipulations. Each observer is a row in the figure, with the first three columns of panels showing the data conditioned by the exposure duration of the stimulus. The last column of panels for each observer is a display of the residuals between observed data $P_{\text{obs}}(\text{CW})$ and predicted performance $P_{\text{pred}}(\text{CW})$, with the dashed line representing an exact correspondence between predicted and observed behavior. Array size 1 = ■, 2 = ●, 3 = ▲, 4 = ◆.

stimulus exposures. They used the diffusion model to decompose the response times into decision and nondecision components, where the latter includes both predecisional perceptual and memory components and postdecisional response selection and execution components. The estimated nondecision times were significantly shorter for single-item displays, consistent with the idea that when the display contains only a single item, the decision process is initiated prior to the report cue. But even on single-item trials, the spread of response times ranged from longer than 500 ms to almost 2,000 ms. Importantly, the display size effect in that task followed the sample-size relationship, implying that there was no difference in the kinds of representations that drive the decision process for single-item and multiple-item displays, although decisions about the latter took significantly longer. The picture that emerges from that study as well as the ones discussed in the previous paragraph is that decisions about brief, masked stimuli are comparatively slow and variable and that much of the processing involved in making them occurs after the perceptual representation of the display has been suppressed by the mask. These results strongly imply that the separable strength and tuning functions that we have reported characterize visual, predecisional stimulus representations rather than postdecisional categorical ones.

Relationship to continuous report work

As indicated in the Introduction, the relationship between changes in reproduction errors observed in a continuous report task and the underlying constraints in memory can be somewhat difficult to determine: Both item- and feature-level constraints on information determine the variability seen in the errors in reproducing feature values, in addition to any systematic patterns of error in responding across different experimental conditions. This problem is reflected in earlier discussions relating to the efficiency of the method of adjustment: As an example, Cornsweet (1962) noted that the method of adjustment made it difficult to know “the process by which [an observer] decides what [response] value to settle for,” (p. 488) particularly when compared to the more standard discrete-choice procedures, which were superseded by the more comprehensive theory of response thresholds provided by signal detection theory.

The extent to which continuous report data can give an estimate of the internal representation in some direct way, as was hoped by Wilken and Ma (2004), is contingent on the relationship between the representation and the subsequent response (i.e., the decision process). One complexity of comparing our results to previous work using a continuous report paradigm is

the lack of a well-developed, widely accepted model of the decision process in continuous report tasks. One candidate model—Smith’s circular diffusion model (2016)—allows a direct translation of the current two-choice model into a form that would be suited for a continuous report task.²

In the standard diffusion model, the drift rate represents the average quality of evidence accumulation toward one of the two response boundaries, with a zero drift rate representing no discriminating evidence provided by underlying perceptual or memory representation, leading to a decision process where the outcome is determined by chance and any response bias (represented as the initial conditions of the decision process). As in the Smith and Sewell model (2013) already described, the sign of the drift rate represents whether the underlying perceptual evidence favors one response outcome over another. An extension of the two-choice diffusion model of Ratcliff (1978), the circular diffusion model uses a drift vector rather than a scalar drift rate to represent the quality and identity of stimulus information. In the circular diffusion model, stimulus identity must also be encoded directly by the decision process: Not only is evidence quantified in terms of the strength of overall evidence, but also the identity information of the stimulus itself. The phase angle of the drift vector represents the identity of the stimulus retrieved from memory (in this case, the stimulus orientation), and the amplitude of the drift vector represents the signal-to-noise ratio of the memory trace. By fixing the phase angle of the drift vector to the actual orientation of the target stimulus (indicating that the observer’s perceptual representation is, on average, veridical) and scaling the amplitude of the vector by $1/\sqrt{m}$ to express the decrease in the signal-to-noise ratio due to the sample-size constraint (see Figure 5), we can integrate the hitting probabilities across orientations for the correct and incorrect response sets to obtain estimated observer accuracy. This integration of the hitting boundary segments a continuously valued outcome from the decision process (which might be observed in a continuous report paradigm) to a discrete response outcome observed in the current two-alternative forced-choice task. The relevant equations for the circular diffusion model are presented in Appendix A.

The result, seen in Figure 6, bears a close qualitative correspondence to the current data. It implies that a statistical decision model developed to characterize decision making in a continuous report task (e.g., by accounting for changes in the distribution of reproduction errors with experimental manipulations) can also characterize the kind of two-choice discrimination performance that is usually modeled using signal detection theory. The model allows a clear distinction between the separable effects of channel tuning and of

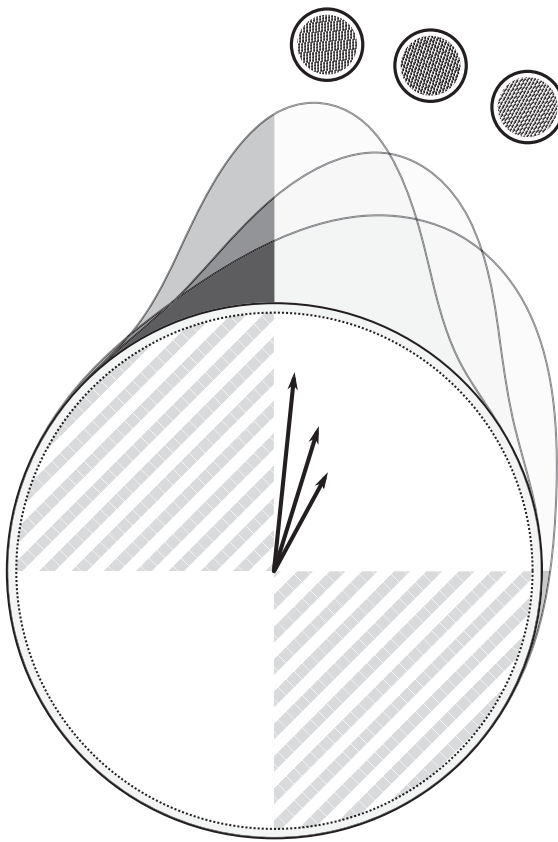


Figure 5. A schematic of how the absorbing boundary of the circular diffusion model may be partitioned into “correct” (the unshaded portion of the circle) and “incorrect” (the shaded portion of the circle) responses to derive discrete choice probabilities. In this figure, the three vectors emanating from the origin of the circle represent the drift vectors for three stimuli, successively presented, of different orientations. The phase angle of the vector represents the identity of the stimulus, and the amplitude of the vector represents the quality of the representation (the signal-to-noise ratio). As a Gabor patch has twofold rotational symmetry, the circle is divided into four segments. The hitting probabilities (von Mises distributions shown on the boundary of the circle) are then integrated to provide the response proportions.

the signal-to-noise ratio with a direct correspondence to the two (polar) dimensions of the drift vector: Channel tuning corresponds to the phase angle of the drift vector and the signal-to-noise ratio corresponds to its length or norm. Independently varying phase and norm produces the separable effects of memory set size and angular separation shown in Figures 5 and 6.

Although this is a qualitative demonstration, it provides a link between standard two-alternative forced-choice tasks of the kind examined in this article and the continuous report paradigm which is a central part of much of the contemporary work on VSTM.

The effect of stimulus exposure duration

One notable difference between the current data and previous data (e.g., Sewell et al., 2014) is in the relationship between stimulus exposure duration and observer sensitivity. The strongest form of the sample-size model of VSTM formation assumes that samples are recruited at a constant rate. This model predicts a linear increase in *squared* sensitivity with stimulus exposure duration, consistent with the data previously reported. We considered models that allowed either sensitivity or squared sensitivity to increase at a linear rate; the linear increase in sensitivity fit substantially better.

Although a linear increase in squared sensitivity is not a necessary prediction of a sample-size model of capacity constraints, it is a strong indicator that the memory system has a fixed rate of information processing, in both the consolidation and maintenance of information. One explanation for the discrepancy between the current fine orientation discrimination work and previous orthogonal orientation discrimination work is that high spatial frequency detectors required to make fine orientation judgements have a slower temporal response than low spatial frequency detectors (Enroth-Cugell & Robson, 1966; Kulikowski & Tolhurst, 1973; Smith, 1995). This would mean that the sensory information required to form a memory representation would be accessible over a smaller fraction of the total stimulus exposure time (see Figure 7). Without additional data to estimate the time in which effective stimulus information can be extracted by the observer from the memory array, the current model must rely upon the total actual display time of the stimulus as a proxy. Future research is required to further constrain both the period in which information can be extracted from the stimulus display and the functional form of that information accumulation.

Conclusion

The results presented in this article provide, we hope, compelling evidence that VSTM performance in a fine orientation discrimination task can be well described by a model in which the tuning of the item representations and their overall strength have separable effects on memory performance. Increasing memory load decreased item memory strength by a proportion predicted by the sample-size model, but item tuning remained invariant. These results are strongly consistent with a model in which memory load affects the strength of individual item representations but not their variability.

The model described comprises both item-level constraints on the total amount of information that can

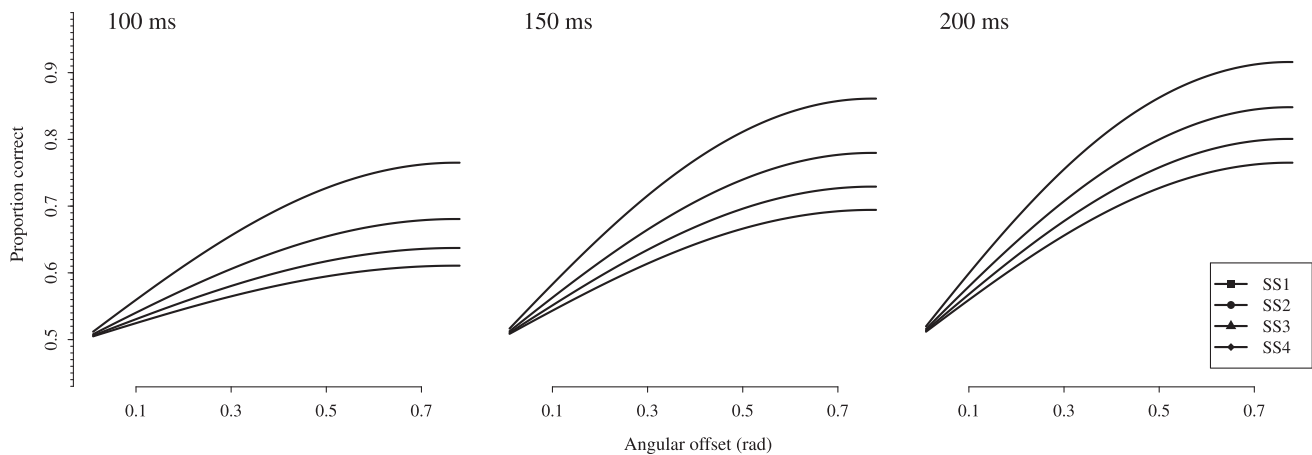


Figure 6. Predicted observer performance in a two-alternative forced-choice fine orientation discrimination task using Smith's (2016) circular diffusion model. Discrete choices were formed by segmenting the circular absorbing boundary into correct and incorrect responses. The phase angle of the drift vector was set to the presented stimulus, and the amplitude of the drift vector was attenuated as a function of the square root of the display size (the sample-size constraint) and as a linear function of the stimulus exposure duration. For illustration purposes we have plotted a slightly larger range of angular offset orientations compared to Figure 3.

be stored in the memory system and feature-level constraints in the ability to distinguish between two similar but different visual items. On an architectural level, these constraints could be obtained by a memory system that is composed of a limited pool of feature-specific detectors, similar to the orientation-specific neural populations found in primary visual cortex (Daugman, 1980; Hubel & Wiesel, 1959, 1962), which are distributed—evenly, in the case of the basic sample-

size model—between the items to be remembered (as depicted in Figure 1). Similar results with other visual features would be required to demonstrate the generality of such a model, but the current data, paradigm, and model could provide some purchase on deeper architectural questions.

Keywords: visual short-term memory, working memory, visual perception, mathematical models of memory, decision making

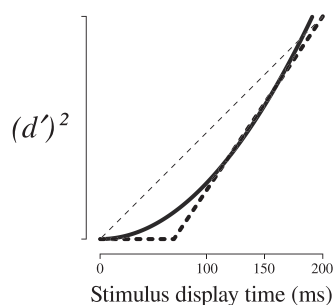


Figure 7. A strong interpretation of the sample-size model predicts a linear increase in squared sensitivity (depicted as a thin dashed line) rather than a quadratic increase as was shown in the current data (depicted as the thick solid line). The discrepancy between the current data and theory (as well as past data) may be explained by the fact that stimulus exposure duration may not provide an adequate measure during which stimulus information is being extracted. If the response of high-spatial-frequency detectors is slow—with a period of negligible information extraction followed by a linear increase, as depicted by the thick dashed line—then the increase in information may be approximated by a quadratic increase due to the assumption that information is accumulated from the physical onset of the stimulus display.

Acknowledgments

This work was supported by an Australian Postgraduate Award to SDL, Australian Research Council Discovery Grant DP140102970 to PLS, and Australian Research Council Discovery Early Career Researcher Award DE140100772 to DKS.

Commercial relationships: none.

Corresponding author: Simon D. Lilburn.

Email: lilburns@unimelb.edu.au.

Address: Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Australia.

Footnotes

¹ The BIC is often written in the form $-2LL + k\ln(N)$, where LL is the maximum log likelihood from the fit of the model to the data. The G^2 is equal to twice the difference between the maximum achievable likeli-

hood, obtained if the predictions of the model and the data coincide exactly, and the maximized likelihood of the model. Because the first of these terms is independent of the model, it will be constant across all models for a given set of data. The BIC is an interval-scale measure, which is unaffected by the addition or subtraction of a constant to or from the likelihoods of all the models being compared. Consequently, the two forms of the BIC are functionally equivalent. We prefer the G^2 form because it has a true zero, at which the data and the model are in perfect agreement, and therefore gives a useful index of the absolute fit of the model. It also has a useful theoretical interpretation as the Kullback–Leibler divergence (the directed distance) between the data and a model. For an example of the BIC defined in terms of G^2 , see Agresti (2003).

² The representation of the continuous report task as a diffusion process is a natural one: Cardozo (1965) provided a statistical model for the method of adjustment that represents the evolution of observer behavior as being that of the Fokker–Planck equation in discrete time.

References

- Adam, K. C., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79–97, <https://doi.org/10.1016/j.cogpsych.2017.07.001>.
- Agresti, A. (2003). *Categorical data analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111, <https://doi.org/10.1111/j.0963-7214.2004.01502006.x>.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628, <https://doi.org/10.1111/j.1467-9280.2007.01949.x>.
- Barton, B., Ester, E. F., & Awh, E. (2009). Discrete resource allocation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1359–1367, <http://doi.org/10.1037/a0015792>.
- Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *The Journal of Neuroscience*, *34*(10), 3632–3645, <https://doi.org/10.1523/JNEUROSCI.3204-13.2014>.
- Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10):7, 1–11, <https://doi.org/10.1167/9.10.7>. [PubMed] [Article]
- Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, *11*(10):6, 1–15, <https://doi.org/10.1167/11.10.6>. [PubMed] [Article]
- Bays, P. M., & Husain, M. (2008, August 8). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854, <https://doi.org/10.1126/science.1158023>.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392, <https://doi.org/10.1177/0956797610397956>.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*(4), 487–502, <http://doi.org/10.1037/a0016797>.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5):4, 1–34, <https://doi.org/10.1167/11.5.4>. [PubMed] [Article]
- Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2011). A neural theory of visual attention and short-term memory (NTVA). *Neuropsychologia*, *49*(6), 1446–1457, <https://doi.org/10.1016/j.neuropsychologia.2010.12.006>.
- Cardozo, B. L. (1965). Adjusting the method of adjustment: SD vs DL. *The Journal of the Acoustical Society of America*, *37*(5), 786–792, <https://doi.org/10.1121/1.1909439>.
- Cornsweet, T. N. (1962). The staircase-method in psychophysics. *The American Journal of Psychology*, *75*(3), 485–491, <https://doi.org/10.2307/1419876>.
- Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, *20*(10), 847–856, [https://doi.org/10.1016/0042-6989\(80\)90065-6](https://doi.org/10.1016/0042-6989(80)90065-6).
- Donkin, C., Kary, A., Tahir, F., & Taylor, R. (2016). Resources masquerading as slots: Flexible allocation of visual working memory. *Cognitive Psychology*, *85*, 30–42, <https://doi.org/10.1016/j.cogpsych.2016.01.002>.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experi-*

- mental Psychology: General*, 113(4), 501–517, <https://doi.org/10.1037/0096-3445.113.4.501>.
- Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, 123(2), 161–177, <https://doi.org/10.1037/0096-3445.123.2.161>.
- Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, 12(6), 1127–1133, <https://doi.org/10.3758/BF03206454>.
- Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology*, 187(3), 517–552, <https://doi.org/10.1113/jphysiol.1966.sp008107>.
- Ester, E. F., Serences, J. T., & Awh, E. (2009). Spatially global representations in human primary visual cortex during working memory maintenance. *The Journal of Neuroscience*, 29(48), 15258–15265, <https://doi.org/10.1523/JNEUROSCI.4388-09.2009>.
- Graham, N. V. S. (1985). Detection and identification of near-threshold visual patterns. *Journal of the Optical Society of America A*, 2(9), 1468–1482, <https://doi.org/10.1364/JOSAA.2.001468>.
- Graham, N. V. S. (1989). *Visual pattern analyzers*. New York: Oxford University Press.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3), 574–591. Retrieved from <https://physoc.onlinelibrary.wiley.com/doi/epdf/10.1113/jphysiol.1959.sp006308>
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1), 106–154, <https://doi.org/10.1113/jphysiol.1962.sp006837>.
- Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *The Journal of Physiology*, 232(1), 149–162, <https://doi.org/10.1113/jphysiol.1973.sp010261>.
- Lee, D., & Chun, M. M. (2001). What are the units of visual short-term memory, objects or spatial locations? *Perception & Psychophysics*, 63(2), 253–257, <https://doi.org/10.3758/BF03194466>.
- Lu, Z.-L., & Doshier, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, 38(9), 1183–1198, [https://doi.org/10.1016/S0042-6989\(97\)00273-3](https://doi.org/10.1016/S0042-6989(97)00273-3).
- Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281, <https://doi.org/10.1038/36846>.
- Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology*, 14(9), 744–751, <https://doi.org/10.1016/j.cub.2004.04.028>.
- Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, 124(1), 21–59, <https://doi.org/10.1037/rev0000044>.
- Orhan, A. E., & Jacobs, R. A. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychological Review*, 120(2), 297–328, <http://dx.doi.org.ezp.lib.unimelb.edu.au/10.1037/a0031541>.
- Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 16(2), 332–350.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, 16(2), 283–290, <https://doi.org/10.3758/BF03203943>.
- Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: 1. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1), 261–282.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108, <http://doi.org/10.1037/0033-295X.85.2.59>.
- Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 127–140. Retrieved from <http://psycnet.apa.org/journals/xhp/26/1/127/>
- Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General*, 139(1), 70–94.
- Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, 61(2), 168–185, <https://doi.org/10.1016/j.neuron.2009.01.002>.
- Rouder, J. N., Morey, R. D., Cowan, N., Zwillig, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National*

- Academy of Sciences, USA*, 105(16), 5975–5979, <https://doi.org/10.1073/pnas.0711295105>.
- Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2014). An information capacity limitation of visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 40(6), 2214–2242, <https://doi.org/10.1037/a0037744>.
- Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2016). Object selection costs in visual working memory: A diffusion model analysis of the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(11), 1673–1693, <https://doi.org/10.1037/a0040213>.
- Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, 102(3), 567–593, <https://doi.org/10.1037/0033-295X.102.3.567>.
- Smith, P. L. (2015). 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, 41–52, <https://doi.org/10.1016/j.jmp.2015.03.007>.
- Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, 123(4), 425–451, <https://doi.org/10.1037/rev0000023>.
- Smith, P. L., Corbett, E. A., Lilburn, S. D., & Kyllingsbæk, S. (2018). The power law of visual working memory characterizes attention engagement. *Psychological Review*, 125(3), 435–451.
- Smith, P. L., Ellis, R., Sewell, D. K., & Wolfgang, B. J. (2010). Cued detection with compound integration-interruption masks reveals multiple attentional mechanisms. *Journal of Vision*, 10(5):3, 1–28, <https://doi.org/10.1167/10.5.3>. [PubMed] [Article]
- Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, 116(2), 283–317, <http://dx.doi.org.ezp.lib.unimelb.edu.au/10.1037/a0015156>.
- Smith, P. L., Ratcliff, R., & Sewell, D. K. (2014). Modeling perceptual discrimination in dynamic noise: Time-changed diffusion and release from inhibition. *Journal of Mathematical Psychology*, 59, 95–113, <https://doi.org/10.1016/j.jmp.2013.05.007>.
- Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Research*, 44(12), 1297–1320, <https://doi.org/10.1016/j.visres.2004.01.002>.
- Smith, P. L., & Sewell, D. K. (2013). A competitive interaction theory of attentional selection and decision making in brief, multielement displays. *Psychological Review*, 120(3), 589–627, <http://doi.org/10.1037/a0033140>.
- Smith, P. L., Sewell, D. K., & Lilburn, S. D. (2015). From shunting inhibition to dynamic normalization: Attentional selection and decision-making in brief visual displays. *Vision Research*, 116, 219–240, <https://doi.org/10.1016/j.visres.2014.11.001>.
- Spitzer, H., Desimone, R., & Moran, J. (1988, April 15). Increased attention enhances both behavioral and neuronal performance. *Science*, 240(4850), 338–340, <https://doi.org/10.1126/science.3353728>.
- Thomas, J. P., & Gille, J. (1979). Bandwidths of orientation channels in human vision. *Journal of the Optical Society of America*, 69(5), 652–660, <https://doi.org/10.1364/JOSA.69.000652>.
- Treue, S., & Martinez-Trujillo, J. C. (1999, June 10). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399(6736), 575–579, <https://doi.org/10.1038/21176>.
- van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, 121(1), 124–149, <https://doi.org/10.1037/a0035234>.
- van den Berg, R., & Ma, W. J. (2014). “Plateau”-related summary statistics are uninformative for comparing working memory models. *Attention, Perception, & Psychophysics*, 76(7), 2117–2135, <https://doi.org/10.3758/s13414-013-0618-7>.
- van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences, USA*, 109(22), 8780–8785, <https://doi.org/10.1073/pnas.1117465109>.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 92–114, <https://doi.org/10.1037/0096-1523.27.1.92>.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 32(6), 1436–1451, <https://doi.org/10.1037/0096-1523.32.6.1436>.
- Vogels, R. (1990). Population coding of stimulus orientation by striate cortical cells. *Biological Cybernetics*, 64(1), 25–31, <https://doi.org/10.1007/BF00203627>.
- Watson, A. B., & Robson, J. G. (1981). Discrimination at threshold: Labelled detectors in human vision.

Vision Research, 21(7), 1115–1122, [https://doi.org/10.1016/0042-6989\(81\)90014-6](https://doi.org/10.1016/0042-6989(81)90014-6).

- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12):11, 1120–1135, <https://doi.org/10.1167/4.12.11>. [PubMed] [Article]
- Woodman, G. F., & Vecera, S. P. (2011). The cost of accessing an object's feature stored in visual working memory. *Visual Cognition*, 19(1), 1–12, <https://doi.org/10.1080/13506285.2010.521140>.
- Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, 10(1), 80–87, <https://doi.org/10.3758/BF03196470>.
- Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology* (rev. ed.). London: Methuen.
- Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation*, 10(2), 403–430, <https://doi.org/10.1162/089976698300017818>.
- Zhang, W., & Luck, S. J. (2008, May 8). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233–235, <https://doi.org/10.1038/nature06860>.

Appendix A: Circular diffusion model

The circular diffusion model of Smith (2016) provides an account of decision making in continuous report tasks, extending the standard diffusion model of Ratcliff (1978). Like the standard diffusion model, the circular diffusion model assumes that responses are the outcome of the sequential sampling of underlying stimulus or memory information. Although the model provides the joint distribution of response times and response angles, these components can be separated, and for this article we are interested only in the distribution of response angles.

The distribution of responses angles was shown by Smith to be equal to the von Mises distribution used to describe the distribution of responses in many existing approaches to the continuous report task. The probability of a response angle θ , thought of as the hitting

point of a diffusion process within a two-dimensional circle of radius a (known as the criterion radius), is

$$P(\theta) = K_{\|\mu\|} \exp\left[\frac{1}{\sigma^2} (a\mu_1 \cos \theta + a\mu_2 \sin \theta)\right].$$

Here σ is the infinitesimal standard deviation of the diffusion process (the square root of the diffusion coefficient), and $\|\mu\|$ is the norm of the drift vector, with μ_1 and μ_2 as the magnitude of the first and second (Cartesian) components of the drift vector. The normalizing constant of the equation is equal to $K_{\|\mu\|} = [2\pi I_0(a\|\mu\|/\sigma^2)]^{-1}$, where $I_0(\cdot)$ is a modified Bessel function of the first kind of order zero.

The standard parameterization of the von Mises distribution is in terms of a location parameter ϕ and a concentration or precision parameter κ , which is related to the norm of the drift vector, the criterion radius, and the diffusion coefficient of the circular diffusion process: $\kappa = a\|\mu\|/\sigma^2$. The drift components of the response-angle density can be reparameterized in terms of the location parameter as $\arctan(\mu_2/\mu_1)$. Substituting these reparameterized terms into the response-angle density, and simplifying using trigonometric identities, one obtains

$$P(\theta) = K_{\|\mu\|} \exp[\kappa \cos(\theta - \phi)] \\ = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \phi)],$$

which is the density function of the von Mises distribution. Further details of the derivation of these equations, as well as of the response-time distribution of the circular diffusion process and applications of the circular diffusion model, can be found in Smith (2016).

For the demonstration in this article, the location of the response-angle distribution was set to the actual stimulus identity and the dispersion scaled as a function of the sample-size relationship (the drift norm decreasing as a function of the square root of the number of items displayed). To produce Figure 6, we set σ to unity, the criterion radius a to 5, and $\|\mu\| = 2.7/\sqrt{m} \cdot \tau$ (where m is the array size and τ is the stimulus exposure duration). The distribution was segmented into four parts, as shown in Figure 5, and the probability density for the correct responses (i.e., those in the quadrant in which the drift vector was pointing or the opposing quadrant) was integrated to get the proportion of correct responses.