

# A detection theory account of change detection

**Patrick Wilken**

Division of Biology, California Institute of Technology,  
Pasadena, CA, USA



**Wei Ji Ma**

Division of Biology, California Institute of Technology,  
Pasadena, CA, USA



Previous studies have suggested that visual short-term memory (VSTM) has a storage limit of approximately four items. However, the type of high-threshold (HT) model used to derive this estimate is based on a number of assumptions that have been criticized in other experimental paradigms (e.g., visual search). Here we report findings from nine experiments in which VSTM for color, spatial frequency, and orientation was modeled using a signal detection theory (SDT) approach. In Experiments 1-6, two arrays composed of multiple stimulus elements were presented for 100 ms with a 1500 ms ISI. Observers were asked to report in a yes/no fashion whether there was any difference between the first and second arrays, and to rate their confidence in their response on a 1-4 scale. In Experiments 1-3, only one stimulus element difference could occur ( $T = 1$ ) while set size was varied. In Experiments 4-6, set size was fixed while the number of stimuli that might change was varied ( $T = 1, 2, 3$ , and 4). Three general models were tested against the receiver operating characteristics generated by the six experiments. In addition to the HT model, two SDT models were tried: one assuming summation of signals prior to a decision, the other using a max rule. In Experiments 7-9, observers were asked to directly report the relevant feature attribute of a stimulus presented 1500 ms previously, from an array of varying set size. Overall, the results suggest that observers encode stimuli independently and in parallel, and that performance is limited by internal noise, which is a function of set size.

**Keywords:** feature judgment, visual short-term memory (VSTM), signal detection theory, change blindness, high-threshold theory, capacity limitations

## Introduction

A critical aspect of any creature's ability to function effectively within a changing environment is the facility to efficiently utilize information from a variety of sensory sources in both its present and its immediate past. The high evolutionary value of such information is implied by the ability of human observers to store various perceptual dimensions, such as spatial frequency, orientation, and hue, with a high degree of fidelity and stability over extended periods of time (Magnussen & Greenlee, 1992; Magnussen, Greenlee, Asplund, & Dyrnes, 1991; Magnussen, Greenlee, & Thomas, 1996; Regan, 1985). It has been shown, for instance, that observers are readily able to detect spatial frequency changes for time periods of upwards of 60 s that are smaller than the Nyquist frequency associated with the spacing between adjacent cones on the fovea (Magnussen, Greenlee, Asplund, & Dyrnes, 1990; Regan, 1985).

In a typical *visual short-term memory* (VSTM) experiment, observers are presented with two displays, each display composed of a number of spatially distinct stimuli. The two arrays are separated by a short temporal interval, usually greater than 80 ms to avoid attentional capture (Kanai & Verstraten, 2004). Observers are asked to decide whether the two arrays were composed of identical stimuli. Performance is believed to be a function of the nature and extent to which a memory of the first display is formed. It is commonly found that an increase in the number of ele-

ments present leads to a monotonic decrease in the sensitivity of observers to differences between the two displays; although for experiments employing suprathreshold stimuli, this decrease is typically only observed after set size has reached around three to four elements (Luck & Vogel, 1997; Pashler, 1988; Vogel, Woodman, & Luck, 2001).

A prominent class of VSTM model proposes that the performance decline associated with increasing set size is caused by a fundamental limit of the number of items that can be encoded, either because the capacity of VSTM itself is limited (Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988; Vogel et al., 2001), or because of a bottleneck in the number of items that can be attended to during the encoding process (Rensink, 2000).

This type of model assumes that VSTM is restricted in storage capacity to only a few items,  $C$  (often estimated to lie in the range of 4 to 5), within a set size  $N$  (Pashler, 1988). The probability that a suprathreshold change will be reported ( $H$ ) is then

$$H = 1 \quad \text{for } N < C;$$

$$H = \frac{C}{N} + \left(1 - \frac{C}{N}\right)F \quad \text{for } N \geq C, \quad (1)$$

where  $F$  is the probability on a given trial that an observer will incorrectly guess "change" when no change has occurred. This model classically envisages VSTM as a single high-level store within which items, often conceived as

bundles of perceptual features, are stored. This type of *high-threshold (HT) model* has been shown to be based on a number of unattractive assumptions (Macmillan & Creelman, 1991), and its applicability has been questioned for other experimental paradigms, such as visual search (Eckstein, Thomas, Palmer, & Shimozaki, 2000; Palmer, 1995; Palmer, Verghese, & Pavel, 2000). First, it posits that the relevant information used to determine whether a visual object has been seen previously is encoded as a discrete unit within the brain (i.e., the item is simply present or not, in the absence of internal noise). Second, it assumes that the internal state engendered in the absence of an item can never contribute to an “item seen” response (i.e., distracters can never produce enough misleading evidence to elicit a false positive answer). Detection theory accounts, which suggest a continuous model of perceptual encoding, are more natural from a cortical viewpoint (Verghese, 2001). In these models, each encoded item is internally represented by a continuous variable, usually with added Gaussian noise, that is positioned within a multidimensional perceptual space.

The purpose of the present study was to determine how well a detection theory framework could account for the basic finding of a performance decline as a function of set size reported within the VSTM literature, and whether this type of model offered alternative insights into the processes leading to this limitation. Our findings suggest that theoretical models that conjecture a high-level storage bottleneck, such as working memory, are unnecessarily complex, and that the assumption of neuronal noise, in conjunction with a simple decision rule, provides a more satisfactory framework to understand VSTM than current HT accounts.

## Layout of study

This report summarizes the results of nine experiments. These experiments fall naturally within three groups; each group is composed of three independent experiments in which the critical feature space manipulated was color, orientation, or spatial frequency.

The first group involves a standard set-size manipulation used in many VSTM experiments (e.g., Luck & Vogel, 1997; Vogel et al., 2001). These results confirmed that our paradigm produces results consistent with those reported previously in the literature. In addition to the standard analysis, confidence ratings were collected, allowing us to generate *receiver operating characteristics* (ROCs). Compared with the current HT account, detection theory models provided a much better match between theoretical predictions and the observed ROCs. This analysis suggested the key factor limiting performance was an increase in noise in the encoded representation of each stimulus element with increasing set size.

The second group of experiments kept set size constant, but systematically varied target number, allowing us

to systematically vary performance while keeping encoding noise constant. This allowed a reduction in the necessary parameter space needed to model the empirical ROCs, and thus a stricter evaluation of the detection theory account.

The third group of experiments attempted to measure the encoding noise in a more direct manner. In these experiments, rather than requiring observers to report in a yes/no manner whether a change had occurred, observers were asked to manipulate a scalar probe to match the relevant feature property of a previously memorized item in an array of variable set size. These experiments, independent of a detection theory analysis, confirmed the increase in noise in the representations of encoded items as a function of set size.

## Methods

### Participants

Fifteen observers participated in each experiment reported in this study. Participants were all students or staff at the California Institute of Technology, aged between 18 and 35 years, with normal or corrected-to-normal acuity, and by self-report normal color vision.

### Apparatus

Stimuli were generated in Matlab, using Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997), and were presented using an Apple G4 computer, with a 128-bit AGP Graphics card, running Mac OS 9.2.2. They were presented on a 21-inch Apple Studio Design monitor, with a resolution of 1024 x 768 pixels, and a refresh rate of 120 Hz.

### Experiments 1-6

#### Stimuli

**Color experiments.** In the color trials, each display was composed of arrays of square stimuli. Each stimulus had a diameter of 1.5°. The color of each stimulus was selected randomly from a palette of seven colors. The CIE (1976) coordinates for the seven colors used for the stimuli were (L, a, b) red (57.2, 79.7, 62.8), blue (85.4, -87.2, 78.4), green (85.4, -87.2, 78.4), cyan (91.1, -41.6, -31.6), yellow (97.2, -16.1, 91.4), purple (66.6, 97.0, -68.8), and black (2.1, -4.4, -3.8). In Experiment 4, an additional color, orange (61.7, 68.4, 65.7), was added to the test palette. The stimuli were presented on a grey (34.4, 10.7, -1.1) background. Each color was selected on the basis of being highly discriminable from the rest of the elements of the palette used, based on the results of pilot tests not reported here.

**Orientation and spatial frequency experiments.** The orientation and spatial frequency stimulus elements used were Gabor patches. The phase of each Gabor, both within and across arrays, was randomized. The SD of the Gaussian envelope used was 11 pixels, and had a peak amplitude of  $32.3 \text{ cd/m}^2$ .

The orientation stimuli had a wavelength of 16 pixels/cycle, which was equivalent to  $0.2^\circ$ . In the orientation experiments, the initial orientation of each stimulus element was assigned at random. If a change occurred to a stimulus element, there was an equal probability of its orientation being either incremented or decremented by an angle of  $\pi/4$  or  $\pi/2$ .

In the spatial frequency experiments, the orientation of each stimulus element was  $0^\circ$  (i.e., vertical). Each stimulus element had an equal probability of being composed of one of three possible wavelengths (8, 16, or 32 pixels/cycle). If a change occurred, there was an equal probability of the change stimulus adopting either of the two remaining spatial frequency values.

**Placement of stimuli.** The head position of observers was unfixed, and viewing distance was on average 66 cm from the display. Stimuli were presented at  $N$  equally spaced points, around an imaginary circle of diameter  $14.4^\circ$  (see Figure 1 for a schematic of the placement of the stimuli). The number of locations,  $N$ , within the circle was equal to the maximum possible number of stimuli that could be present within an array from a particular experiment (i.e.,  $N = 8$  for color and spatial frequency and  $N = 6$  and 5 for orientation, for the set-size and target-number experiments, respectively). If, in Experiments 1-3, less than the maximum number of possible stimuli was present within an array (i.e., set size within a trial was less than  $N$ ),

the stimuli were placed randomly on a subset of the possible stimulus locations. The minimum spacing between adjacent stimuli (i.e., when  $N = 8$ ) was  $6.5^\circ$ . All stimuli were displayed on a dark grey background ( $2.3 \text{ cd/m}^2$ ).

### Procedure

In Experiments 1-3, set size was varied ( $N = 2, 4, 6$ , and 8 for color and spatial frequency and  $N = 2, 3, 4$ , and 5 for orientation), while number of targets ( $T = 1$ ) was kept constant. In Experiments 4-6, set size was kept constant ( $N = 8$  for color and spatial frequency and  $N = 6$  for orientation), while the number of possible targets was systematically varied ( $T = 1, 2, 3$ , and 4).

Data were collected from participants over two sessions. Each session consisted of five blocks, and each block was composed of 128 trials. The order of trials within all blocks was counterbalanced using an ABBA design. In Experiments 1-3, trials were counterbalanced for set size and change/no-change trials; In Experiments 4-6, trials were counterbalanced for target number and change/no-change trials. The position of each stimulus was randomly assigned on a trial-by-trial basis.

Participants were seated in a dimly lit room. Trials began with the presentation of a fixation point, in the form of a small cross ( $0.2^\circ \times 0.2^\circ$ ), placed at the center of the display. At 250 ms after the onset of the fixation cross, the first array of stimuli was presented for 100 ms. After the offset of the first array, the screen remained blank for 1500 ms. Immediately thereafter a second array was presented for 100 ms. Overall there was a probability of .5 that within a particular trial the two arrays would be composed of stimuli identically matched in the appropriate feature dimension. In Experiments 4-6, on any change trial, there was an equal probability that the change would involve one, two, three, or four stimuli.

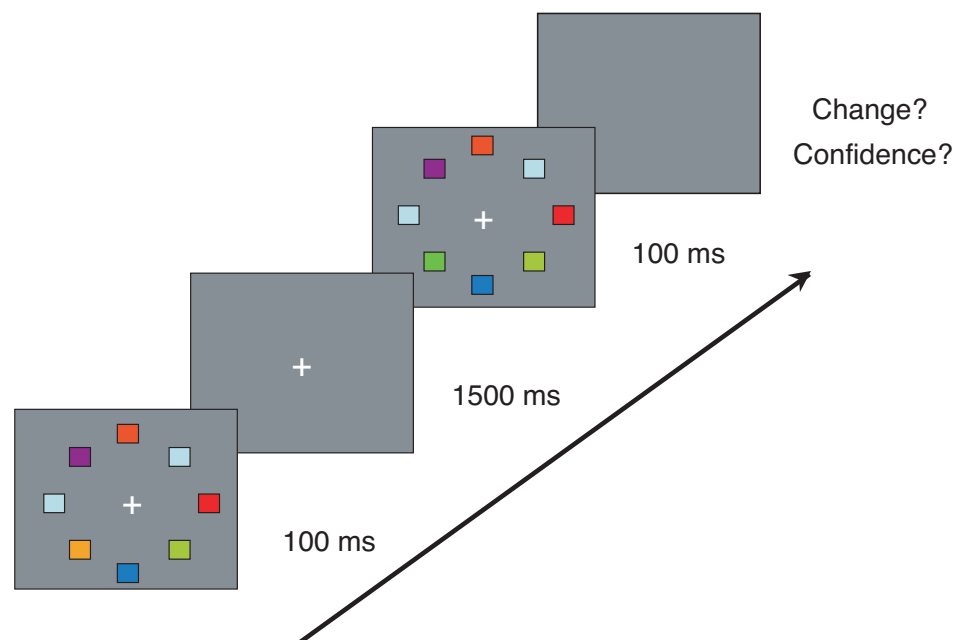


Figure 1. Schematic of a single trial used for the set size and the target number experiments. In the orientation and spatial frequency experiments, the colored squares were replaced with Gabor patches.

Before data collection began, participants were informed that their task was to determine whether the two arrays within a trial consisted of elements identical in the appropriate feature dimension. They were instructed to press the “8” key of the keyboard if they detected a change, and the “9” key if they did not detect a change. After pressing either the “8” or “9” keys, they were instructed to indicate their confidence in their response by pressing either the “1”, “2”, “3,” or “4” keys (where “1” indicated a very high confidence in their response, “2” that they were somewhat confident, “3” somewhat unconfident, and “4” very unconfident). If an observer incorrectly indicated the presence or absence of a change, an auditory tone was sounded immediately after they had indicated their confidence. The importance of accuracy, rather than speed of response, was emphasized to all participants.

## Experiments 7-9

Unless otherwise stated, the methodology was the same as that used in Experiments 1-3.

### Stimuli

**Color experiment.** A palette of 252 colors was used. The CLUT values were assigned to ensure all the presented colors were highly saturated. The CLUT value  $n$  was assigned the value:

$$255 * [1 - n/84 \quad n/84 \quad 0] \text{ for } 0 \leq n \leq 84$$

$$255 * [0 \quad 2 - n/84 \quad n/84 - 1] \text{ for } 85 \leq n \leq 168$$

$$255 * [n/84 - 2 \quad 0 \quad 3 - n/84] \text{ for } 169 \leq n \leq 252$$

Fractional values were rounded to the next highest whole number. The probe stimulus consisted of a color

wheel consisting of the 252 possible color values. The wheel was an annulus with an outer diameter of  $3.0^\circ$  and an inner diameter of  $2.1^\circ$ .

**Orientation experiment.** The stimuli used in the orientation judgment experiment were identical to the Gabor elements used in the previous orientation experiments. The probe stimulus was assigned one of 30 possible orientations, equally spaced between 0 and  $2\pi$  deg.

**Spatial frequency experiment.** In the spatial frequency judgment experiment, each stimulus element was randomly assigned one of 16 spatial frequency values, between 12 pixels/cycle and 24 pixels/cycle ( $.2^\circ$  and  $.4^\circ$ ). The spatial frequency values were spaced in a linear fashion such that adjacent values were equidistant. In all other aspects, the stimulus elements were the same as those used in the two previous spatial frequency experiments.

The probe stimulus was randomly assigned one of 30 values, equally spaced in a linear fashion, such that the middle 16 values were identical to those of the stimulus palette. In all other respects, the properties of the probe matched those of the stimulus elements.

### Procedure

As in Experiments 1-3, set size was varied ( $N = 2, 4, 6$ , or 8 for color and spatial frequency and  $N = 2, 3, 4$ , and 5 for orientation) (see Figure 2).

Data were collected from participants over a single session. Each session consisted of 10 blocks, and each block was composed of 64 trials. The order of trials within all blocks was counterbalanced for set size using an ABBA design. The feature of interest for the particular experiment (i.e., color, orientation, or spatial frequency), as well as the position of the cue and the stimuli, was randomly assigned on a trial-by-trial basis.

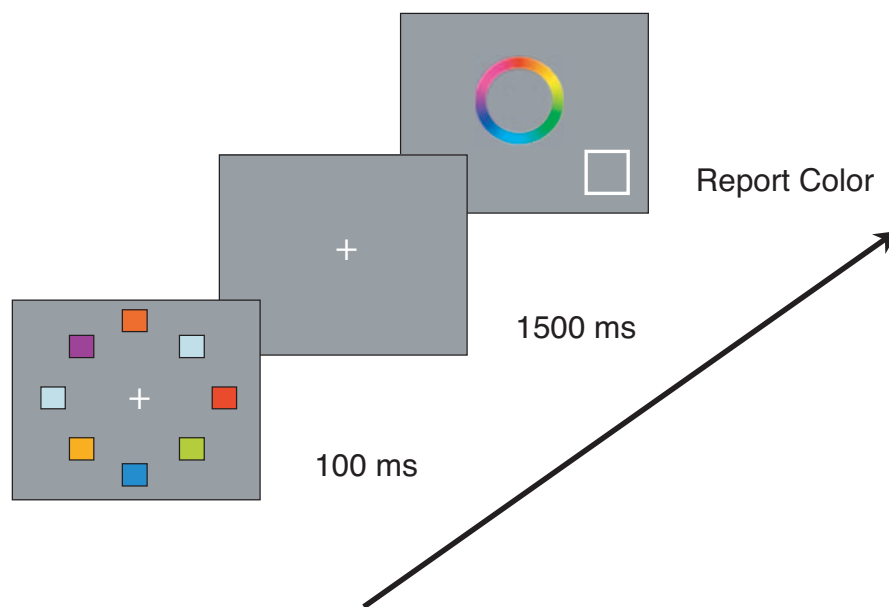


Figure 2. A schematic timeline for a single trial in the color judgment experiment. In the orientation and spatial frequency experiments, the colored squares were replaced by Gabors, and the color wheel was replaced by a probe Gabor.



Trials began with the presentation of a fixation point, in the form of a small cross ( $0.2^\circ \times 0.2^\circ$ ), placed at the center of the display. At 250 ms after the onset of the fixation cross, the array of stimuli was presented for 100 ms. After the offset of the first array, the screen remained blank for 1500 ms. Immediately thereafter a square cue ( $3^\circ \times 3^\circ$ , composed of lines 1 pixel wide) was centered around the location of one of the previously presented items. At the same time, a test probe was displayed centrally. Both the test probe and the cue remained present until the trial was finished.

Before data collection began, participants were informed that their task was to match as closely as possible the relevant stimulus property of the probe stimulus with that of the stimulus at the cued location. In the color experiment, they were asked to indicate the cued stimulus property by using the mouse to point and click at the part of the color-wheel that most closely matched the feature property of the cued stimulus. In the orientation experiment, participants were instructed to match the cued feature property by using the right and left arrows on the keyboard to rotate the probe Gabor (clockwise and counterclockwise, respectively). In the spatial frequency experiment, participants also used the arrow keys on the keyboard to change the spatial frequency of the probe stimulus (the right arrow increasing the spatial frequency, left arrow decreasing it). In the orientation and spatial frequency experiments, observers recorded their responses by pressing the space bar on the keyboard. As in previous experiments, the importance of accuracy, rather than speed of response, was emphasized to all participants.

## Detection theory models of change

Depending on the nature of the decision process, detection theories can be divided into either *first-order* or *second-order integration* models (Shaw, 1982). In first-order integration accounts, the relevant information from each stimulus is pooled, and a single decision is made on that combined information. Conversely, in second-order models, a separate decision is made on each relevant stimulus attribute, and a final response is based on the assimilation of these decisions (Palmer et al., 2000). Here we consider two general accounts of VSTM based on prototypical examples of first-order or second-order integration models. Full technical details of the detection theory and HT models used can be found in [Appendix A](#).

## Maximum absolute differences model

The *maximum absolute differences* model (MAD) is a detection theory generalization of an HT account previously reported by one of the authors (Wilken, 2001), and is based closely on Shaw's (1980) *independent decisions* model. In this class of second-order integration model (Palmer,

1990; Shaw, 1980), an observer attempting to detect a change among  $N$  elements is considered to be monitoring  $N$  noisy channels, each channel representing information associated with a single item within the display. A change in an item will be reflected in a change in the signal of one of the channels. If the alteration in this signal is greater than a particular threshold, a change will be reported. It is further assumed that decisions across channels are independent, and hence  $N$  independent decisions are made about  $N$  elements within a display. Because the signal in each channel is noisy, there is a certain probability that the signal in a non-change channel will pass above the detection threshold and a change will be reported erroneously.

Here we assume that it is the absolute size of this change that is important for the detection of change, and not its sign (e.g., a change from red to green will be as readily detected as a change from green to red). Such a differencing operation is common in describing classical same-different tasks (Macmillan & Creelman, 1991), and is the simplest operation of its kind. This assumption of the MAD model leads to a somewhat different mathematical exposition from the max model variant (Eckstein et al., 2000; Palmer, 1990, 1995; Palmer et al., 2000; Shaw, 1980) previously employed to explain performance in visual search tasks.

Moreover, we assume a fixed effective distance in stimulus representation space (and thus sensitivity) between different stimuli at one location in the two displays. It should be noted that by doing this, we collapse all represented magnitudes of individual changes (e.g., a transition from red to green versus one from red to orange); however, we do expect their differences to translate into differences in performance.

Within the MAD model, the typical finding that an increase in set size leads to a monotonic decline in performance in VSTM tasks can be attributed to two potential causes. First, increases in the number of channels being monitored will lead to an increase in the likelihood that noise within a non-change channel will be mistakenly attributed to an actual change (i.e., decisional noise). Second, independent of decisional noise, increases in set size may amplify the noise present within each channel. For instance, the assumption that observers are limited by a fixed number of samples they can extract from a visual scene,  $N$ , leads to a prediction from central limit theorem that noise within each (equally) monitored channel will increase proportional to  $N$ , because the number of samples per channel will then be proportional to  $1/N$  (Palmer, 1990). Naturally, other assumptions will lead to different changes in internal noise as a function of  $N$ . Here we take an agnostic viewpoint, and retain a single free parameter as our noise estimate within a single monitored channel. In cases in which change involves a single target element, the MAD model approximates the performance of an ideal observer (Palmer et al., 2000).

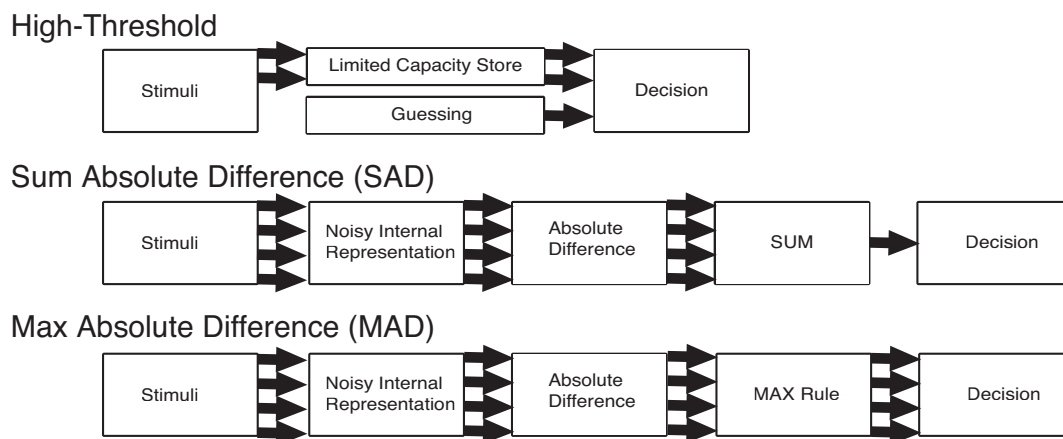


Figure 3. A diagram of the information flow in the HT, SAD, and MAD models.

### Sum of absolute differences model

In addition to the MAD model, a *sum of absolute differences* model (SAD) was assessed. This first-order integration account shares the MAD model's assumptions that each stimulus attribute is represented by a noisy internal state, and that a stimulus change is represented by an absolute difference between matched states in the first and second displays. It differs from the MAD model in that it presumes that the absolute differences calculated for each stimulus are summated to form the distribution upon which a single decision is based.

Figure 3 offers a schematic of the information flow in the three models examined in this study.

## Change detection as a function of set size

The first three experiments performed a set-size manipulation frequently employed in previous VSTM experiments (Vogel et al., 2001). This experimental manipulation allowed us to assess whether our experimental methodology produced results broadly consistent with experimental paradigms that have applied an HT account (Luck & Vogel, 1997). However, unlike these previous experiments, confidence ratings were also collected allowing us to generate empirical ROCs that could then be compared with theoretical predictions from the MAD and SAD models of change detection.

### Results

As shown in Figure 4, there was a general decline in the ability of observers to detect a change between the first and second arrays, with hit rates monotonically decreasing, and false alarm rates monotonically increasing as a function of set size.

An analysis of the color change data using an HT model (Pashler, 1988) implies a capacity estimate  $C$  of  $3.80 \pm .13$  ( $M \pm SE$ ), consistent with previous studies that have

suggested visual working memory for color has a storage capacity of three-to-five items (Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988; Wilken, 2001). An HT analysis for the orientation and spatial frequency data yielded a slightly lower estimate for  $C$  ( $2.58 \pm .27$ ,  $2.51 \pm .18$ , for orientation and spatial frequency, respectively), but one still in general accord with the broadly agreed upon storage limit of VSTM.

The confidence-interval data were pooled over observers, and empirical ROCs were obtained for all set sizes. For each model tested, the best-fitting parameters were determined through a multidimensional error-minimization procedure. At every iteration of this minimization, two steps were executed. First, the locations of the criteria for all set sizes were estimated using the empirical response frequencies in the no-change condition and the noise distribution from the model with the parameters values at the current iteration. Next, for each set size, a chi-squared statistic was computed between the average numbers of responses for a single observer in each response category in the change condition and the numbers expected from the model signal distribution with the current parameter values and the estimated criterion locations. These chi-squared values were summed over set sizes, and their sum was minimized using an algorithm provided by Matlab (The Mathworks, Inc.) based on the Nelder-Mead simplex (direct search) method. To correct for problems associated with pooling across nonlinear data sets, and to obtain error bars for the parameter estimates, a jackknife procedure was employed (Quenouille, 1949; Tukey, 1958). The goodness of the resulting fit was measured by the above-mentioned total chi-squared statistic.

We first attempted to fit the ROC data using the standard HT model described in Equation 1. As can be seen in Figure 4, the empirical ROCs were regular in shape, quite unlike the theoretical straight lines predicted by HT accounts (Swets, 1986a, 1986b). Not surprisingly, as can be seen in Table 1, the resulting HT fit was very poor.

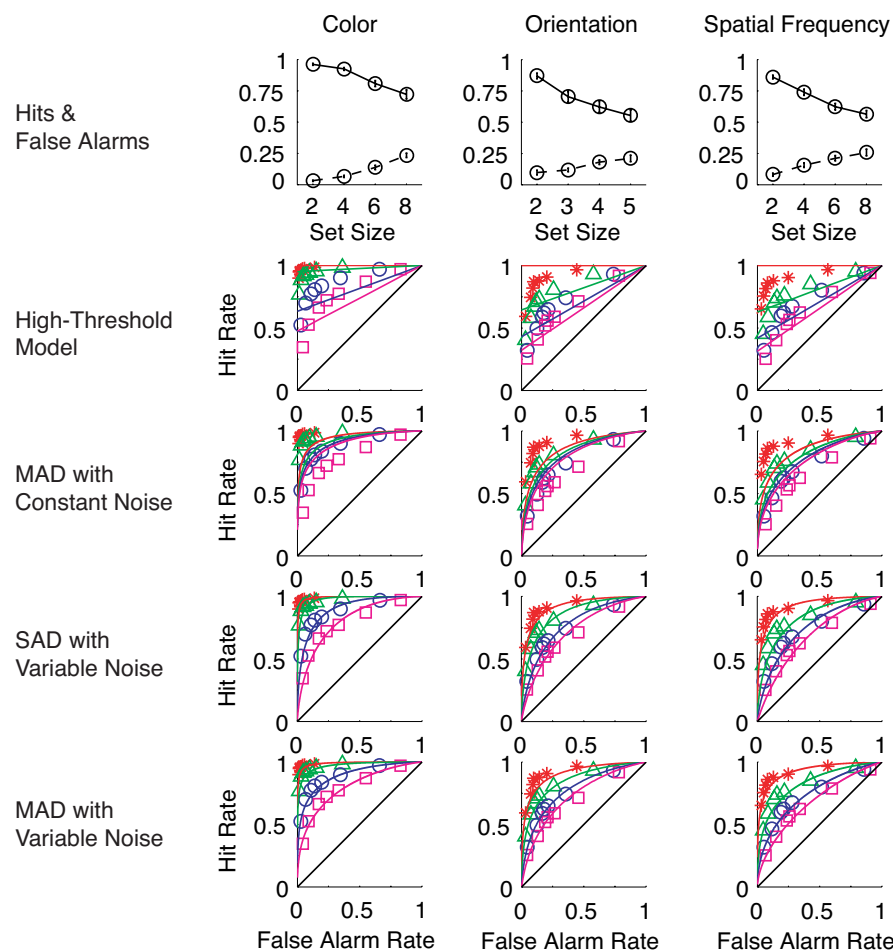


Figure 4. Raw data plus model fits for the color, orientation, and spatial frequency set-size experiments. Top row shows overall hit rates (solid line) and false alarm rates (dashed line) for 15 observers. In all cases the *SE* bars fall within the circular symbols. The lower four rows show example model fits for the HT model, MAD with constant noise and equal variance, and SAD and MAD with the variable noise and equal variance assumptions. Different symbols represent performance at different set sizes: set-size 2 for color, orientation and SF – red stars; set-size 4 for color and SF, set-size 3 for orientation – green triangles; set-size 6 for color and SF, set-size 4 for orientation – blue circles; set-size 8 for color and SF, set-size 5 for orientation – purple squares.

Type of model	Constant noise	Equal variance	Color $\chi^2$ (df)	Orientation $\chi^2$ (df)	SF $\chi^2$ (df)
HT	-	-	>2503 (27)	>305 (27)	>310 (27)
MAD	Yes	Yes	120 (27)	50.5 (26)	81.4 (27)
	Yes	No	119 (26)	48.2 (25)	79.5 (26)
	No	Yes	32.8 (24)	28.2 (23)	36.0 (24)
	No	No	26.6 (20)	22.6 (19)	34.3 (20)
SAD	Yes	Yes	94.4 (27)	53.4 (26)	85.5 (27)
	Yes	No	43.0 (26)	52.0 (25)	54.6 (26)
	No	Yes	40.6 (24)	31.0 (23)	60.8 (24)
	No	No	21.5 (20)	29.6 (19)	27.2 (20)

Table 1. Results of parameter fitting for HT, MAD, and SAD models for the variable set-size experiments.

SDT accounts offer a rich theoretical basis with which to understand the effects of set size on performance in VSTM tasks. We assessed how well decisional noise alone could account for changes in performance with set size (our constant-noise condition), and whether allowing noise to increase with set size generated a significantly better fit for the observed ROCs (our variable-noise condition). In addition, we assessed whether allowing the variances of the change and the no-change distributions to differ provided a significantly better fit to the observed distributions, as previous studies have shown that the assumption of equal variance between signal and noise distributions is often violated (Swets, 1986a, 1986b).

The ROCs in our SDT models are determined by the effective perceptual distance,  $d$ , between different stimuli at one location in the two displays (unlike the usual sensitivity parameter  $d'$ , this does not assume equal and unitary variances), and the amount of noise in their representations. For color and spatial frequency, all SDT models have a redundant scaling parameter, which we use to fix  $d$ . In our models,  $d$  is assumed to be constant across changes in set size. This leaves our constant-noise model with a single free parameter, the SD of the noise, and for the variable-noise models four parameters, the SDs of the noise associated with each set size. However, for orientation, because the stimulus space is circular, there is no such scaling, and we always estimate the effective distance  $d$  together with the parameter(s) of the noise (see Appendix A for a discussion

of the measurement of noise in von Mises distributions). Thus, the SDT models in the orientation paradigm have one degree of freedom less than those in the color and spatial frequency paradigms.

Before discussing the relative effectiveness of various SDT models to account for the ROC data, it is important to note that all SDT accounts offered better fits to the ROC data than the standard HT model. Moreover, this result cannot be solely attributed to a larger parameter space used by the SDT accounts. Even when the SDT accounts shared the same degrees of freedom as the HT accounts, the fit for the SDT models was much better ( $\chi^2$  equal to 2503 vs. 120 and 94, for HT vs. MAD and SAD for the color experiment;  $\chi^2$  equal to 305 vs. 50.5 and 53.4 for HT vs. MAD and SAD for the spatial frequency experiment).

Next, within-model comparisons were performed to assess whether the reduction in the  $\chi^2$  values associated with rejecting the constant noise assumption could be explained purely as a result of the increase in the parameter space (see Wickens, 2002, for an explanation of this technique). This analysis showed that even when taking the increase in the size of the parameter space into account, for both MAD and SAD, the model fit was significantly better when the assumption of constant noise was relaxed ( $p < .001$  for all comparisons, no correction for multiple comparisons). Figure 5 shows the change in the noise parameter for the MAD model. In all three feature modalities, the increase in

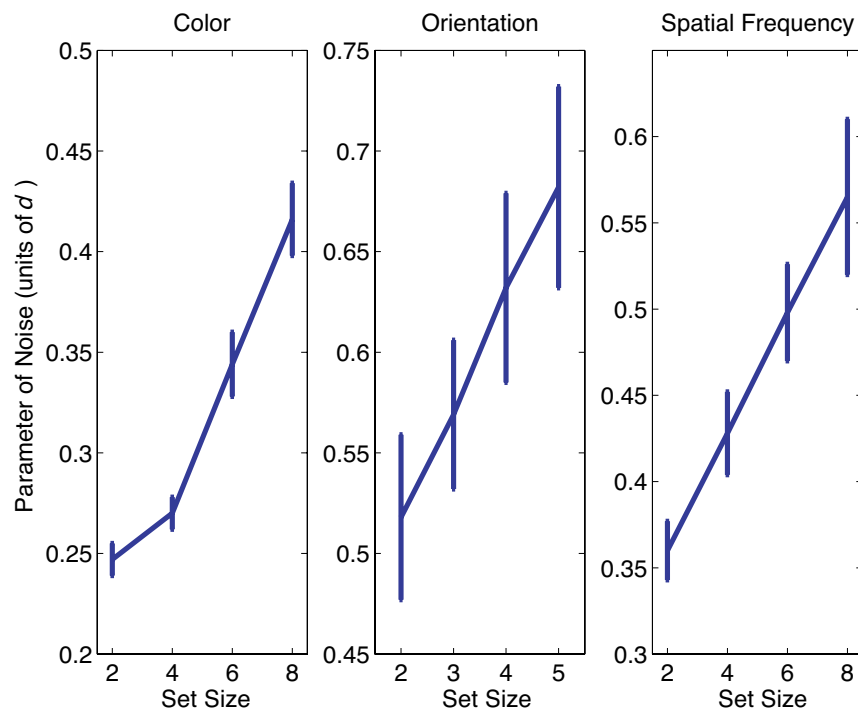


Figure 5. Behavior of the noise estimated from the MAD equal-variance model as a function of set size. For color and spatial frequency, the plot shows the SD of the noise in units of the effective distance  $d$  described in the text. For orientation, the plot shows the equivalent parameter  $s$  (see Appendix A, Equation 5) in units of the estimated  $d$ .



noise is monotonic and roughly linear. This strongly suggests that as more items are encoded, the encoding of each item becomes noisier.

A similar evaluation was performed to assess the relative improvement in  $\chi^2$  associated with relaxing the equal variance assumption. In this case, no significant change in  $\chi^2$  was found for the MAD fits for any of the experiments, or for the SAD fit for orientation ( $p > .05$ ), whereas a significant improvement was found for SAD fits for color and spatial frequency ( $p < .001$  for both comparisons).

## Change detection as a function of target number

The results of the set-size experiments supported the value of a detection theory approach for modeling performance in VSTM tasks, with both the SAD and MAD models providing good fits to the experimental data. These experiments suggest that the noise of the internal percept increases monotonically as a function of set size.

Given these results, we decided to perform a second set

of experiments in which set size was fixed ( $N = 8$  for color and spatial frequency and  $N = 6$  for orientation), but performance systematically varied by changing the target number ( $T = 1, 2, 3$ , or  $4$ ). This manipulation had two main advantages: first, it allowed an assessment of the robustness of the SDT approach when performance was varied in quite different manner; and second, by fixing set size (and therefore, presumably, internal noise) it allowed a substantial reduction in the necessary parameter space needed for modeling, in turn allowing a more sensitive comparison between models.

## Results

Performance improved as a function of the number of elements changing, as demonstrated by the monotonically increasing hit rates as a function of target number shown in Figure 6.

Experiments 4-6 were analyzed in a similar manner to Experiments 1-3, with the essential difference that only one noise parameter was used for all four ROCs, because set size was kept constant.

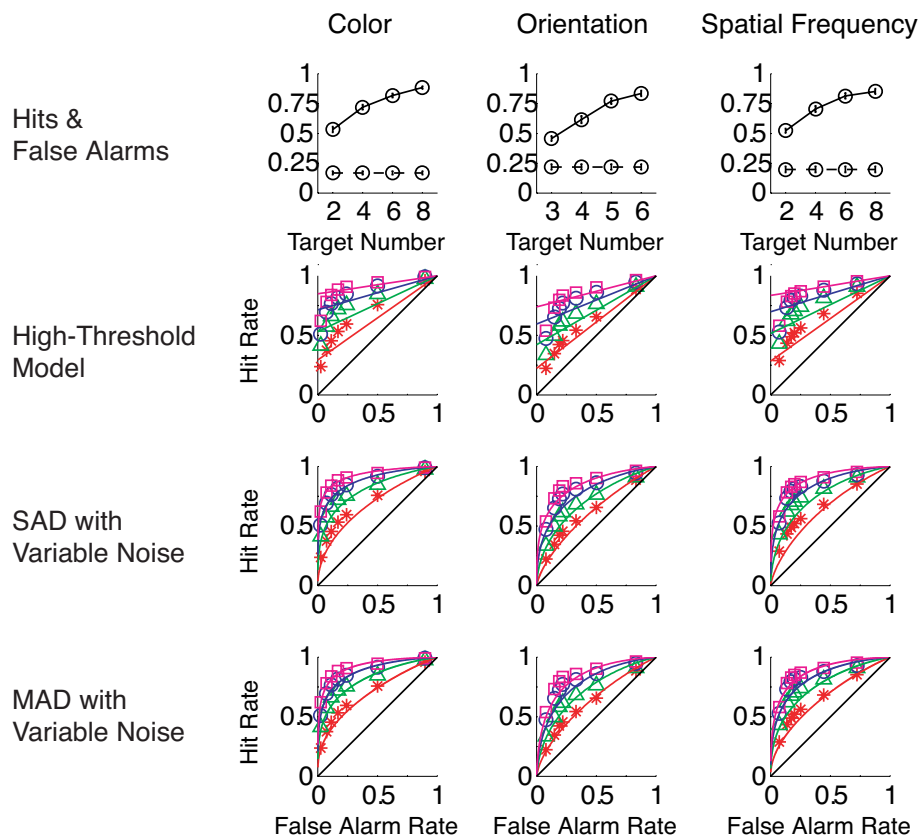


Figure 6. Raw data plus model fits for the color, orientation, and spatial frequency target-number experiments. Top row shows overall hit rates (solid line) and false alarm rates (dashed line) for 15 observers. In all cases the *SE* bars fall within the circular symbols. Experiments generated only one false alarm rate; for comparison purposes, the same false alarm rate is shown for each target number condition. The lower three rows provide example model fits for the HT model, and SAD and MAD with variable noise and unequal variance assumptions. Symbols represent performance in the different target number conditions: one target – red stars; two targets – green triangles; three targets – blue circles; and four targets – purple squares.

An analysis of the color change data was performed, using a modified form of the HT model (Pashler, 1988) to take into account variations in target number, and this model implied capacity estimates of  $C$  of  $2.35 \pm .14$  for color,  $1.79 \pm .17$  for orientation, and  $2.27 \pm .26$  for spatial frequency. These results again are broadly consistent with previous estimates of the storage of visual working memory (Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988; Wilken, 2001).

As can be seen in Figure 6, the empirical ROCs were regular in shape, again quite unlike the straight lines predicted by HT accounts (Swets, 1986b).

An examination of Table 2 makes a number of points immediately apparent. As in the set-size experiments, the fit of the HT model is much worse across all experiments compared to the fits associated with SAD and MAD models. Once more, the MAD model in general offers a better fit for the data than the SAD model (the only exception being the analysis of the orientation experiment, under the equal variance assumption).

Contrasting with the results of the previous set-size experiments, a within-model analysis found that there was a significant reduction in  $\chi^2$  values associated with relaxing the equal variance assumption for all MAD model fits ( $p < .001$  for color and orientation and  $p < .01$  for spatial frequency). Removing the equal variance assumption was also found to significantly improve the SAD model fits for the color ( $p < .001$ ) and orientation ( $p < .05$ ) experiments.

Why was there stronger evidence against the equal variance assumption in the target number experiments, than in the set-size experiments? We can think of at least two reasons. First, by keeping the set size constant, we were able to perform a more sensitive test of the change in  $\chi^2$  values. Perhaps more crucially, it would be expected that if in fact the change and no change distributions were unequal, that this would become more apparent as the number of targets with the display increased.

## An estimation of the internal representation

Explicit within the MAD and SAD models is the assumption that change detection is a process of comparison between stored elements of the first array with corresponding elements within the second array. As such, the second array can be considered a complex probe into the information encoded for each element within the first array. This probe is necessarily complex, requiring comparisons across all elements in the first encoded array and their matching elements within the second probe array, with a corresponding decision rule needed to determine whether stimulus elements differ between the first and second arrays.

The results of the set-size experiments suggest that change becomes harder to detect as a function of increasing

set size primarily as a function of increasing noise within the stored representation of each stimulus element. Given that the detection theory account is broadly correct, it should be possible to directly show, as a function of set size, an increase in noise for the feature properties of a single encoded element, without inferring these indirectly from the responses generated when observers are required to make comparisons across multiple stimulus elements between arrays.

Type of model	Constant noise	Equal variance	Color $\chi^2$ (df)	Orientation $\chi^2$ (df)	SF $\chi^2$ (df)
HT	-	-	979 (27)	322 (27)	898 (27)
MAD	Yes	Yes	66.1 (27)	54.1 (26)	63.4 (27)
	Yes	No	18.9 (26)	28.6 (25)	57.4 (26)
SAD	Yes	Yes	138.6 (27)	35.7 (26)	75.5 (27)
	Yes	No	25.4 (26)	30.7 (25)	74.9 (26)

Table 2. Results of parameter fitting for HT, MAD, and SAD models for variable target number experiments.

The following three judgment experiments were essentially variations on the previous set-size experiments, with the following major modification: rather than using a second array as a probe of change detection, at the time when the second array would have appeared, the location of a single element within the first array was cued, and observers were asked to modify the feature properties of a probe item to match those of the cued element (these experiments were inspired by earlier work by Prinzmetal and colleagues [Prinzmetal, Amiri, Allen, & Edwards, 1998; Prinzmetal, Nwachuku, Bodanski, Blumenfeld, & Shimizu, 1997]). This method greatly simplifies the decision process, and avoids the necessity of observers having to perform multiple comparisons across stored stimulus items. The demonstration of a monotonic increase in noise of a single item as a function of set size in this task would offer strong independent support for the detection model account of VSTM.

## Results

The three judgment experiments show a consistent picture: with increasing set size, the variance of the estimate of the cue feature rises in a monotonic fashion. As shown in Figure 7, this increase in noise shows an approximately linear trend when measured in terms of standard deviation.

It is important to note that HT accounts predict a very different pattern of results: noise should increase little if at all until working memory capacity was full, at which point observer's judgment noise should show a sudden increase as observers start to guess on a certain proportion of trials the relevant stimulus attribute. No such abrupt change in the noise is evident in the data.

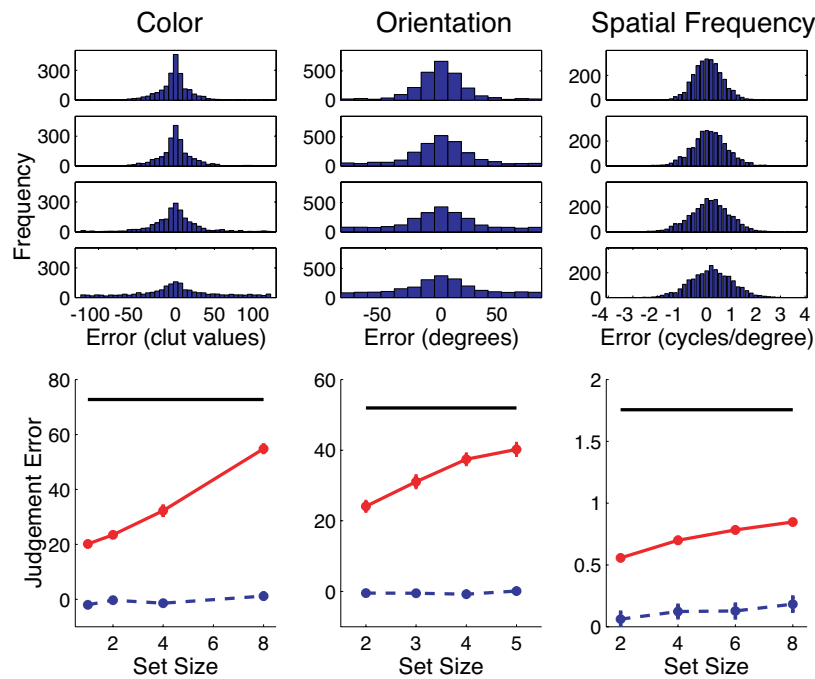


Figure 7. Histograms and summary statistics for judgments of color, orientation, and spatial frequency. The top four rows show response histograms for color, orientation, and spatial frequency. Each histogram summarizes responses for 15 observers at a single set size. Set size increases down each column. The bottom row presents summary statistics: mean of judgment error (dashed blue line), SD of judgment error (solid red line), and chance performance (solid black line). In several cases, SE bars fall within symbols.

The broadly linear increase of the SD of the judgment error with set size was consistent to that observed in the set size experiments, as can be seen by a comparison of the estimated noise of the internal representation shown in Figure 5 with the measured noise of the internal representation shown in Figure 7. These results are consistent with the belief that the major limiting factor in change detection is noise in the internal representation of each encoded item, and not a limitation in the number of encoded items.

### Bias in judgments

Observers showed no systematic bias in their estimates of either color or orientation. Interestingly, observers did show a bias in estimating spatial frequency, systematically reporting higher spatial frequencies as lower than their actual values, and conversely, reporting lower spatial frequencies as higher than their real values; further, as set size increased, this bias systematically increased (see Figure 8). Prinzmetal et al. (1998) reported a somewhat different effect in their judgment experiments in which observers systematically over-reported spatial frequencies. He suggested that this was an example of *contrast overconstancy* in which observers overestimate the spatial frequencies of suprathreshold stimuli (Georgeson, 1991; Georgeson & Sullivan, 1975). However, this explanation cannot account for our data in which observers are also systematically *underestimating* higher spatial frequencies.

The spatial frequency results can best be described as a general bias in reporting toward the mean. A trivial explanation of this result might be expected to hold if observers

were to guess on some proportion of trials. However, an examination of the histograms in Figure 7 shows no obvious sign that observers were guessing in any substantial manner.

A tentative alternative explanation would presume that observers judge spatial frequencies by combining observed values with an internal template based on prior experience, perhaps through a weighted sum in which the weights de-

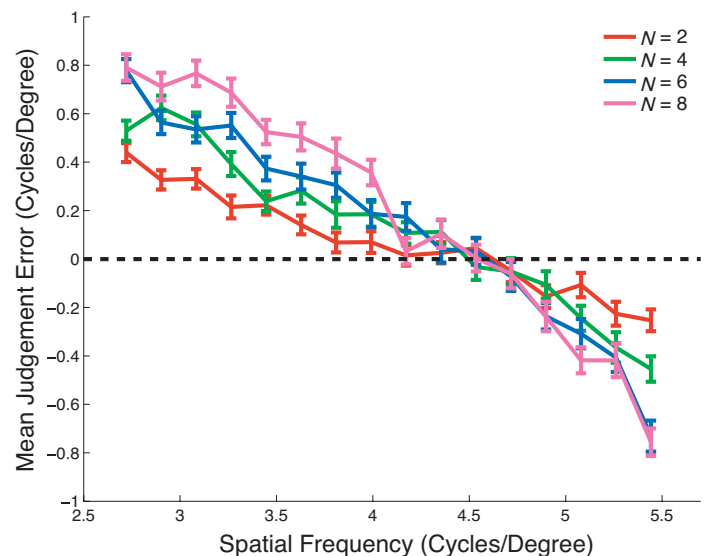


Figure 8. Mean judgment error as a function of the presented spatial frequency as a function of set size. Note greater regression toward the mean with increasing set size.

crease with increasing variance. As set size increases, the variance of the noise in the observations increases, and less weight is given to them. This would increase the magnitude of the shift toward the mean of the template.

## General discussion

Although HT models are largely discredited for simple detection and discrimination, they persist in the literature for both visual search (e.g., Rensink, 2000; Wolfe, 2003) and visual short-term memory (e.g., Alvarez & Cavanagh, 2004; Woodman, Vogel, & Luck, 2001). Recently, a number of authors have criticized the HT explanation of visual search, arguing instead for the advantages offered by a detection theory account (Palmer et al., 2000; Verghese, 2001). This study has attempted to show the analogous advantages offered by a detection theory account to understanding VSTM.

There are good reasons why a detection theory approach is to be preferred over traditional HT accounts. It is neurally implausible that items are encoded in visual memory without noise. While most supporters of HT accounts would probably agree with this, they would presumably also argue that the addition of noise does not change the fundamental fact that VSTM suffers a capacity limit caused by a limited number of slots within a high-level store (see, for instance, Wright, Alston, & Popple, 2002). However, a SDT account of VSTM has little in common with this “slots-plus-noise” sketch. Our analyses suggest that it is unnecessary to postulate a second higher level storage stage to account for the decrease in performance in VSTM tasks with increasing set size. Rather the assumption of neuronal noise, plus a simple decision rule, is sufficient to capture much of the complexity of the VSTM data.

By framing the theoretical account in this alternate manner, a different picture of the underlying processes associated with change detection becomes apparent. Within-model analyses of both the SAD and MAD detection theory accounts consistently show a significantly improved fit for the data when the constant noise assumption is relaxed, even when the resultant increase in the parameter space is taken into account (the correctness of this approach is independently supported by the results for the three judgment experiments). By this account, the postulation of a high-level storage bottleneck is unnecessary, and distracts from the underlying cause of the observed capacity limit (i.e., increasing noise as a function of set size). It remains an interesting empirical question whether the changes in noise associated with set size are due to factors prior to encoding (e.g., saliency and/or attentional effects; Braun, Koch, & Davis, 2001) and/or caused by interference between items encoded within memory (Magnussen & Greenlee, 1997).

At first glance, our finding of an increase in noise with set size is perhaps surprising given the common result that set-size effects in visual search tasks can be modeled with

the assumption of constant noise (see Eckstein et al., 2000; Palmer et al., 2000). It is certainly possible that this increase in noise is due to memory factors irrelevant to measures of performance in visual search tasks. However, it is important to note that detection theory has been unable to parsimoniously explain visual search performance in the presence of heterogeneous distracters. In this case, performance declines have been observed that are greater than would be expected from a detection theory account assuming decisional noise alone (Rosenholtz, 2001), a situation similar to that found in the present change detection experiments.

The main aim in this research has been to show the relative benefits of a detection theory approach for understanding VSTM. Working on the assumption that for a first approach a simple model that fits much of the data is more convincing than a complex model that fits correspondingly more, we chose to develop two very simple detection theory accounts. While the observed fit to the data for both SDT models was much better than the HT account, it was also apparent, in the target number manipulation experiments, that neither model fully captured the structure of the data. It would thus appear that at least one of the underlying assumptions of the MAD and SAD models is wrong. Perhaps the most obvious failure of our approach has been an inability to develop an ideal-observer model for change detection. If human observers are ideal observers, they use a likelihood ratio as a criterion, rather than values of the internal representation itself. For a single detection task, this yields the same ROC, but when one integrates information from multiple items, the ideal observer behaves differently from the MAD and SAD models. For a large number of stimuli with one target, the ideal observer is very similar to the max rule (Green & Swets, 1978; Palmer et al., 2000). In contrast, if all items are targets, the ideal-observer theory makes the same predictions as the sum rule (Green & Swets, 1966). For the general case of  $N$  stimuli and  $T$  targets, the ideal observer cannot be analytically computed (Palmer et al., 2000), while numerical calculations are very difficult. The development of an ideal-observer analysis would be an important next step in the development of a full account of the mechanisms associated with change detection.

An examination of the VSTM literature reveals a theoretical split between those experimenters employing a more traditional psychophysical methodology, using threshold stimuli (e.g., Magnussen, 2000), and those from a high-level vision background utilizing suprathreshold changes (e.g., Olson & Jiang, 2002). While the latter typically envisage a single high-level limited-capacity store, the former often conceptualize VSTM as a series of parallel, special purpose feature stores, occurring post-V1, but prior to mid-level vision (Magnussen & Greenlee, 1999). When these theoretical differences have been acknowledged, it has typically been argued that threshold and suprathreshold changes map onto different memory systems (Vogel et al., 2001).



Because our experimental methodology utilized suprathreshold changes, and our detection theory account is compatible with psychophysical accounts of VSTM, it appears unnecessary to propose a dichotomy between the memory systems probed by these two experimental techniques.

The HT account of memory that is currently popular in the literature makes an appealingly minimal set of assumptions. However, it appears unable to account for various empirical findings reported within this study. The assumption that change detection is made in the presence of neuronal noise, in conjunction with a simple decisional rule, offers a straightforward alternative explanation of performance in these tasks. One conclusion of this approach is that the “four-item limit” commonly reported when using HT models of VSTM does not reflect the number of items held within a high-level store, but rather is an artifact due to mounting noise in internal stimulus representations as set size is increased.

## Appendix A

### Models

Appendix A describes the models discussed. In particular, we present the predictions that each model makes for the ROC curves in our change detection experiments.

#### High-threshold model

In an HT model, there is a limited capacity to store items. The stored items are encoded perfectly, and whether a stored item changed or not can be determined with certainty. Items that fall outside the capacity limit are subject to guessing at a fixed rate. We denote the number of changing items in a change trial with  $T$ . Suppose first that  $T = 1$ . If the capacity is  $C$  and the total number of items presented is  $N$ , the hit rate  $H$  is a linear function of the false alarm rate  $F$  (see Equation 1).

If there are  $T$  targets, there are  $\binom{N}{C}$  ways to store  $C$  items, and  $\binom{N-T}{C}$  ways to store  $C$  non-target items. Therefore, we find the following prediction for the ROC:

$$H = 1 - \frac{\binom{N-T}{C}}{\binom{N}{C}} + \frac{\binom{N-T}{C}}{\binom{N}{C}} F \quad (2)$$

#### Signal detection models

In signal detection theory, the fundamental assumption is that the internal representation of a task-relevant feature (in this study, color, orientation, or spatial frequency) is noisy. We assume that the noise is normally distributed. In

the orientation experiments, the representation space is periodic, and taken to be  $[0, \pi)$ . The noise distribution is a normal distribution on the circle, a so-called Von Mises distribution (Mardia, 1972). Its density is given by

$$p(\theta) = \frac{1}{2\pi I_0\left(\frac{1}{s^2}\right)} e^{\frac{\cos 2(\theta - \theta_0)}{s^2}}. \quad (3)$$

Here,  $\frac{1}{s^2}$  is the concentration parameter and  $\theta_0$  is the circular mean.  $I_0$  is the modified Bessel function of the first kind of order 0 (Abramowitz & Stegun, 1965, pp. 374-377); it serves as a normalization. When  $s \rightarrow \infty$ , the von Mises distribution becomes the uniform distribution on  $[0, \pi)$ . When  $s \rightarrow 0$ , it behaves as a normal distribution on the real line with SD  $s$ .

#### Absolute difference models

We describe the representation of a change as the absolute difference between the representations of the two items. Differencing models are well known in signal detection treatments of same-different designs (see Chapter 6, Macmillan & Creelman, 1991). For color and spatial frequency, we use the absolute difference of two normally distributed variables with SDs and a distance  $d$  between their means. This has the density

$$s_d(x) = \frac{1}{\sigma\sqrt{2\pi}} \left( e^{-\frac{(x-d)^2}{2\sigma^2}} + e^{-\frac{(x+d)^2}{2\sigma^2}} \right), \quad (4)$$

where  $x \in [0, \infty)$ . For orientation, the absolute difference of two Von Mises-distributed variables with an angular distance,  $d$ , between their means is given by the density

$$s_d(x) = \frac{1}{\left(2\pi I_0\left(\frac{1}{s^2}\right)\right)^2} \left( I_0\left(\frac{2\cos(x-d)}{s^2}\right) + I_0\left(\frac{2\cos(x+d)}{s^2}\right) \right). \quad (5)$$

Here,  $x \in \left[0, \frac{\pi}{2}\right)$ . In a single change detection task (one item in the display), the signal is given by  $s_d(x)$  for some  $d > 0$ , whereas the noise is given by  $s_0(x)$ . A “yes” response is made whenever an internal representation  $x$  exceeds a decision criterion  $c$ . The hit rate is:

$$h = P(\text{'yes'} | \text{signal}) = P(x > c | \text{signal}) = 1 - \int_0^c s_d(x) dx \quad (6)$$

and the false alarm rate is:

$$f = P(\text{'yes'} | \text{noise}) = P(x > c | \text{noise}) = 1 - \int_0^c s_0(x) dx. \quad (7)$$



“Unequal variances” in this context means that the two variables that are subtracted in the noise are assumed to have a different variance than the two variables subtracted in the signal. When there are multiple items, an integration rule is needed. We examine the max and sum rules.

### Max absolute difference model

The max rule states that a “yes” response is made if the maximum of the internal representations of all the items exceeds a criterion. This is equivalent to an independent-decisions model (Shaw, 1980), in which at each location an independent decision is made, and the overall response is “yes” if at least one of the locations yields the response “yes.” If there are  $N$  locations and  $T$  targets, this gives rise to the following overall hit and false alarm rates:

$$H = 1 - (1 - h)^T (1 - f)^{N-T} \quad (8)$$

$$F = 1 - (1 - f)^N, \quad (9)$$

where  $f$  and  $h$  are the single change detection hit and false alarm rates above. Examples of MAD distributions of different values of  $N$  and  $T$  are shown in the first column in Figure 9.

### Sum absolute difference model

In the sum absolute difference model, the overall representation is obtained from the sum of the stimulus representations at all locations. If there are  $T$  targets  $S_1, \dots, S_T$  with probability density  $s_d(x)$ , and  $N - T$  non-targets

$S_{T+1}, \dots, S_N$  with probability density  $s_0(x)$ , then the overall representation is the variable

$$S(N, T) = \sum_{i=1}^T S_i + \sum_{j=T+1}^N S_j. \quad (10)$$

A value drawn from this distribution of sums is compared with an overall decision criterion. Hit and false alarm rates can only be computed numerically. Examples of SAD distributions of different values of  $N$  and  $T$  are shown in the second column in Figure 9.

## Acknowledgments

This research was supported by the Rosamund Alcott Fellowship, California Institute of Technology (PW), the Netherlands Organisation for Scientific Research and the Swartz Foundation (WJM), and National Institute of Mental Health and National Science Foundation grants to Christof Koch.

We would like to thank Tamara Becher for running the initial version of Experiments 1-3 as a summer undergraduate research fellow (SURF). We also thank William Banks, Jochen Braun, Jon Driver, Miguel Eckstein, Farshad Moradi, John Palmer, William Prinzmetal, Ron Rensink, Dan Simons, and Preeti Verghese for helpful discussions and useful suggestions. Finally, we would like to thank Christof Koch for both the intellectual and practical support he offered us during our time as postdoctoral scholars within his laboratory.

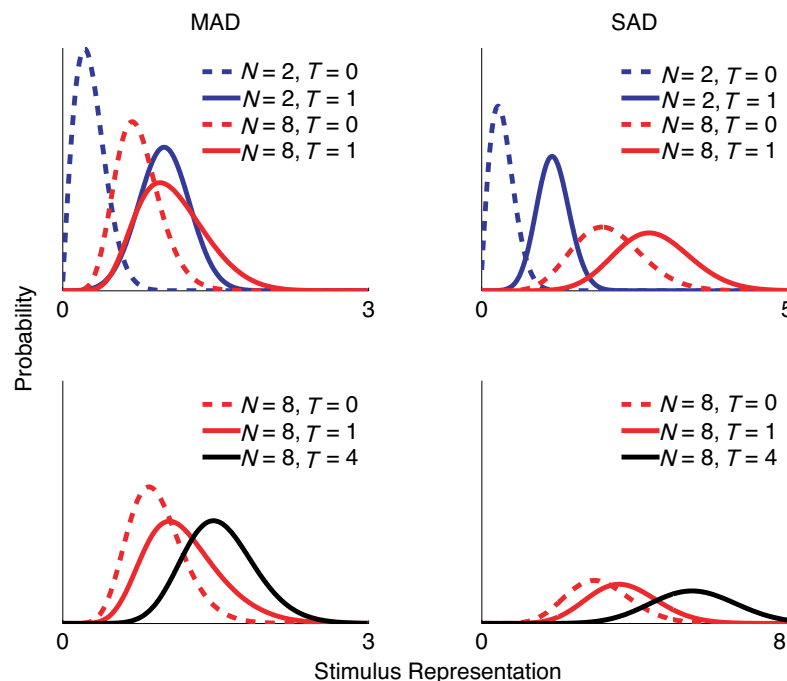


Figure 9. Examples of MAD and SAD equal-variance distributions for color. No-change distributions are in dashed lines and change distributions are in solid lines. The upper row shows them for two set sizes ( $N = 2$  and  $N = 8$ ) with variable noise ( $SD$ s from Experiment 1); the lower row shows them for two target numbers ( $T = 1$  and  $T = 4$ , with  $N = 8$ ) with  $SD$ s from Experiment 4. The scales on the x-axes are in units of  $d$ . The scales on the y-axes are identical across subgraphs.

Commercial relationships: none.

Corresponding author: Patrick Wilken.

Email: patrick.wilken@nat.uni-magdeburg.de.

Address: Otto von Guericke Universität, Fakultät für Naturwissenschaften, Universitätplatz 2, D-39106 Magdeburg, Germany.

## References

- Abramowitz, M., & Stegun, I. A. (1965). *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*. New York: Dover Publications.
- 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. [PubMed]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433-436. [PubMed]
- Braun, J., Koch, C., & Davis, J. L. (2001). *Visual attention and cortical circuits*. Cambridge, MA: MIT Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87-114. [PubMed]
- Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics*, 62(3), 425-451. [PubMed]
- Georgeson, M. A. (1991). Contrast overconstancy. *Journal of the Optical Society of America A*, 8(3), 579-586. [PubMed]
- Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *Journal of Physiology*, 252(3), 627-656. [PubMed]
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Green, D. M., & Swets, J. A. (1978). Detection and recognition. *Psychological Review*, 85, 192-206.
- Kanai, R., & Verstraten, F. A. J. (2004). Visual transients without feature changes are sufficient for the percept of a change. *Vision Research*, 44(19), 2233-2240. [PubMed]
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279-281. [PubMed]
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- Magnussen, S. (2000). Low-level memory processes in vision. *Trends in Neurosciences*, 23(6), 247-251. [PubMed]
- Magnussen, S., & Greenlee, M. W. (1992). Retention and disruption of motion information in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 151-156. [PubMed]
- Magnussen, S., & Greenlee, M. W. (1997). Competition and sharing of processing resources in visual discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 23(6), 1603-1616. [PubMed]
- Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological Research*, 62(2-3), 81-92. [PubMed]
- Magnussen, S., Greenlee, M. W., Asplund, R., & Dyrnes, S. (1990). Perfect short-term memory for periodic patterns. *European Journal of Cognitive Psychology*, 2, 245-262.
- Magnussen, S., Greenlee, M. W., Asplund, R., & Dyrnes, S. (1991). Stimulus-specific mechanisms of visual short-term memory. *Vision Research*, 31(7-8), 1213-1219. [PubMed]
- Magnussen, S., Greenlee, M. W., & Thomas, J. P. (1996). Parallel processing in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 22(1), 202-212. [PubMed]
- Mardia, K. V. (1972). *Statistics of directional data*. London: Academic Press.
- Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception & Psychophysics*, 64(7), 1055-1067. [PubMed]
- 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. [PubMed]
- Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science*, 4(4), 118-123.
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, 40(10-12), 1227-1268. [PubMed]
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, 44(4), 369-378. [PubMed]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437-442. [PubMed]
- Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention. 1. Colour, location, orientation and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 261-282.

- Prinzmetal, W., Nwachuku, I., Bodanski, L., Blumenfeld, L., & Shimizu, N. (1997). The phenomenology of attention. 2. Brightness and contrast. *Consciousness & Cognition*, 6(2-3), 372-412. [PubMed]
- Quenouille, M. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society (Series B)*, 11, 68-84.
- Regan, D. (1985). Storage of spatial-frequency information and spatial-frequency discrimination. *Journal of the Optical Society of America A*, 2(4), 619-621.
- Rensink, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition*, 7(1-3), 345-376.
- Rosenholtz, R. (2001). Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance*, 27(4), 985-999. [PubMed]
- Shaw, M. L. (1980). Identifying attentional and decision-making components in information processing. In R. S. Nickerson (Ed.), *Attention and Performance VIII* (pp. 277-295). Hillsdale, NJ: Lawrence Erlbaum Associated.
- Shaw, M. L. (1982). Attending to multiple sources of information. *Cognitive Psychology*, 14, 353-409.
- Swets, J. A. (1986a). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, 99(2), 181-198. [PubMed]
- Swets, J. A. (1986b). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, 99(1), 100-117. [PubMed]
- Tukey, J. W. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics*, 29, 614.
- Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, 31(4), 523-535. [PubMed]
- 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, 92-114. [PubMed]
- Wickens, T.D. (2002). *Elementary signal detection theory*. Oxford ; New York: Oxford University Press.
- Wilken, P. (2001). *Capacity limits for the detection and identification of change: Implications for models of visual short-term memory*. Unpublished doctoral dissertation, University of Melbourne, Melbourne.
- Wolfe, J.M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences*, 7(2), 70-76. [PubMed]
- Woodman, G.F., Vogel, E.K., & Luck, S.J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science*, 12(3), 219-224. [PubMed]
- Wright, M.J., Alston, L., & Popple, A.V. (2002). Set-size effects for spatial frequency change and discrimination in multiple targets. *Spatial Vision*, 15(2), 157-170. [PubMed]