

## Research Article

# A Dynamic Neural Field Model of Visual Working Memory and Change Detection

Jeffrey S. Johnson,<sup>1</sup> John P. Spencer,<sup>2,3</sup> Steven J. Luck,<sup>4,5</sup> and Gregor Schöner<sup>6</sup>

<sup>1</sup>Department of Psychology, University of Wisconsin–Madison; <sup>2</sup>Department of Psychology, University of Iowa; <sup>3</sup>The Iowa Center for Developmental and Learning Sciences, University of Iowa; <sup>4</sup>Center for Mind and Brain, University of California, Davis; <sup>5</sup>Department of Psychology, University of California, Davis; and <sup>6</sup>Institut für Neuroinformatik, Ruhr-University

**ABSTRACT**—Efficient visually guided behavior depends on the ability to form, retain, and compare visual representations for objects that may be separated in space and time. This ability relies on a short-term form of memory known as visual working memory. Although a considerable body of research has begun to shed light on the neurocognitive systems subserving this form of memory, few theories have addressed these processes in an integrated, neurally plausible framework. We describe a layered neural architecture that implements encoding and maintenance, and links these processes to a plausible comparison process. In addition, the model makes the novel prediction that change detection will be enhanced when metrically similar features are remembered. Results from experiments probing memory for color and for orientation were consistent with this novel prediction. These findings place strong constraints on models addressing the nature of visual working memory and its underlying mechanisms.

Human thought and behavior arise within dynamic and highly complex visual environments. Behaving efficiently within such environments depends on the ability to form, retain, and update visual representations as objects and events change over time. Input to the visual system is not continuous, however; rather, it is frequently interrupted by blinks, eye movements, and other visual disruptions. As a result, detecting changes in ongoing events often depends on the ability to compare visual percepts formed at different points in time. This ability relies on a short-term form of visual memory known as visual working memory

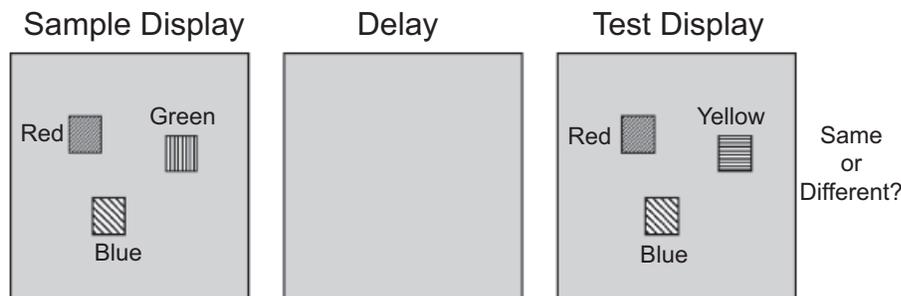
(VWM).<sup>1</sup> Although a considerable body of research has shed light on the neurocognitive systems subserving this form of memory, few theories have addressed how populations of neurons can form, maintain, and compare visual representations. Here, we present a neurally grounded model that integrates these cognitive processes, and we report experiments testing a novel prediction derived from this integration.

Empirical studies of change detection have relied on variants of the simple task shown in Figure 1. Observers view a sample display (e.g., an array of simple objects or an image of a real-world scene), which is followed by a brief disruption of some sort (e.g., an eye movement, a “mud splash,” or a blank screen) and the appearance of a second, test display. The test display either is the same as the sample or differs from it in some way—for instance, the color of one item may have changed (for reviews, see Luck, in press; Rensink, 2002). In the *one-shot* change-detection task shown in Figure 1, a single test display is presented, and observers make an unspeeded response, indicating whether the test display is the same as or different from the sample display (see Luck & Vogel, 1997). In *flicker* change-detection tasks, the original and changed displays alternate, separated by brief blank intervals, until the observer indicates that a change has been detected (see Pashler, 1988; Simons & Rensink, 2005).

Successful change detection in these tasks depends on several factors. First, the information present in the sample display must be accurately perceived and encoded in VWM (Jolicoeur & Dell’Acqua, 1998; Vogel, Woodman, & Luck, 2006). Second, the information must be stably and accurately maintained across the

Address correspondence to Jeffrey S. Johnson, Department of Psychiatry, University of Wisconsin–Madison, 6001 Research Park Blvd., Madison, WI 53719, e-mail: jsjohnson3@wisc.edu.

<sup>1</sup>*Working memory*, as we use the term here, refers to a system or process that actively maintains information over brief temporal intervals for use in service of some task; in our case, the specific task is comparing sensory inputs that are separated by temporal gaps.



**Fig. 1.** Change-detection task used to explore properties of visual working memory for simple features (adapted from Luck & Vogel, 1997). A sample display is followed by a delay and then a test display. The task is to indicate whether the sample and test are the same or different. This illustration shows a task with color stimuli; different fill patterns are used to represent different solid colors.

delay. Third, the visual memory of the sample display must be compared with relevant information in the test display (Mitroff, Simons, & Levin, 2004), and a decision must be generated. The failure to detect changes when they occur, or *change blindness*, can arise when any one of these processes fails.

Contemporary research using the change-detection paradigm has revealed properties of each of the processes involved in visual comparison. For instance, VWM representations are established very rapidly ( $\sim 50$  ms/item; see, e.g., Gegenfurtner & Sperling, 1993; Vogel et al., 2006) and in an all-or-none fashion (Zhang & Luck, 2008). Moreover, only a limited amount of information ( $\sim 3$ – $4$  objects' worth) can be maintained at any given time (Cowan, 2001; Luck & Vogel, 1997). When the amount of information present in the sample display exceeds this capacity, mechanisms of attention play a role in selecting relevant aspects of the display for encoding and maintenance (Hollingworth, Shrock, & Henderson, 2001; Schmidt, Vogel, Woodman, & Luck, 2002). Finally, the detection of changes at test has been found to depend on a process that compares working memory representations with incoming sensory inputs (Mitroff et al., 2004). This process occurs largely in parallel, with detected changes producing an active change signal that elicits rapid orienting to the location of the change (Hyun, Woodman, Vogel, Hollingworth, & Luck, in press).

At another level, research has begun to elucidate the neural systems involved in change detection. Event-related potential and functional imaging studies have shown that the separate components of change detection engage a distributed network of neural populations in the inferior temporal, posterior parietal, and prefrontal cortices (Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002; Todd & Marois, 2004; Vogel & Machizawa, 2004; Xu & Chun, 2006). Additionally, the detection of changes at test has been shown to engage many of the same neural systems implicated in visual selective attention (Beck, Rees, Frith, & Lavie, 2001; Pessoa & Ungerleider, 2004).

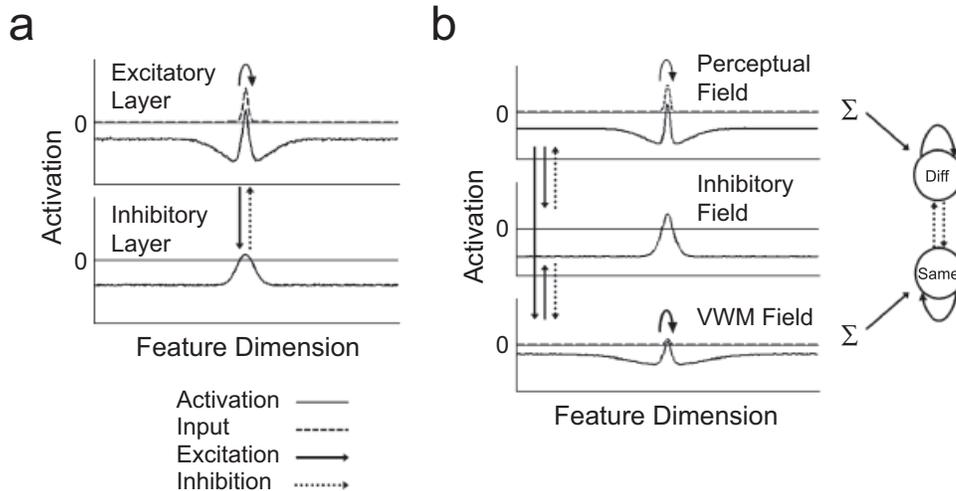
In summary, considerable progress has been made in understanding the component neural and behavioral processes involved in change detection. Although several formal models

have addressed how visual representations are formed and maintained over time (see, e.g., Amit, Bernacchia, & Yakovlev, 2003; Compte, Brunel, Goldman-Rakic, & Wang, 2000), none have addressed the process of comparing multiple visual memory representations with incoming sensory inputs. In this article, we describe a layered neural architecture that implements encoding and maintenance in VWM, and we show how these processes can be linked to a plausible comparison process. The model also makes novel and counterintuitive behavioral predictions, which we tested in a series of experiments.

#### A DYNAMIC NEURAL FIELD MODEL OF VWM AND CHANGE DETECTION

To account for VWM and change detection in a neural framework, we have developed a model that builds on the dynamic field theory (DFT) of visuospatial cognition (Simmering, Schutte, & Spencer, 2008; Spencer, Simmering, Schutte, & Schöner, 2007). The DFT is in a class of continuous-attractor neural network models originally developed to capture the dynamics of neural activation in visual cortex (Amari, 1977; see also Wilson & Cowan, 1972). The generic form of models in this class consists of a single layer of feature-selective excitatory neurons reciprocally coupled to a separate layer of inhibitory interneurons (see Fig. 2a). Locally excitatory and laterally inhibitory interactions within the network allow the formation of localized peaks, or “bumps,” of activation representing, for instance, estimates of specific sensory inputs (e.g., the retinal position or color of a stimulus; see Jancke et al., 1999); in some cases, these peaks may be sustained in the absence of continuing input. Simple networks of this type can realize elementary perceptual and memory processes (see discussion in Grossberg, 1980). However, capturing performance in change-detection tasks also requires specification of the processes underlying visual comparison.

To this end, we have developed the three-layer architecture depicted in Figure 2b. This architecture was inspired by the *canonical cortical circuit* proposed by Douglas and Martin



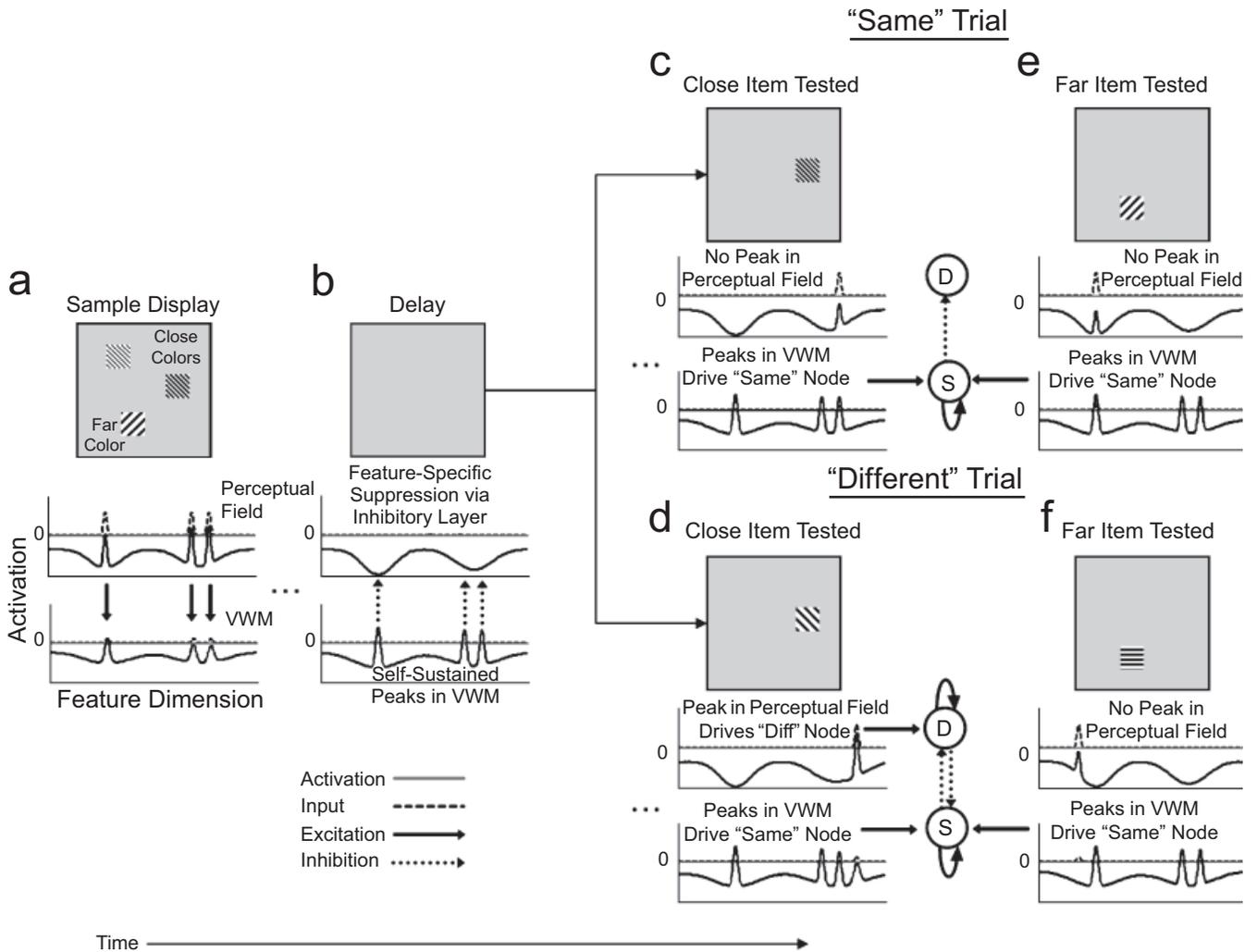
**Fig. 2.** Two- and three-layer dynamic neural field models of visual working memory (VWM). The thin, solid horizontal line in each field marks the activation threshold (conventionally set to be 0), the point at which interactions among neurons within and between layers become engaged. The two-layer model (a) consists of a single population of feature-selective excitatory neurons coupled to a similarly tuned population of inhibitory neurons. This simulation depicts the formation of a peak of activation following localized input to the excitatory layer. Input takes the form of a Gaussian distribution that is centered at a particular field location and has a specified strength and width. Once activation goes above threshold (i.e., 0) in the excitatory layer, activation is passed to the inhibitory layer, which, in turn, passes broad inhibition back to the excitatory layer. Locally excitatory interactions among neurons in the excitatory layer (solid, curved arrow) keep neurons in a highly active state, whereas inhibitory feedback from the inhibitory layer keeps excitation localized by preventing the diffusion of activation throughout the field. The three-layer model (b) contains two populations of excitatory neurons (perceptual and VWM fields) reciprocally coupled to a single population of inhibitory neurons (inhibitory field). Input is applied to both excitatory fields, but input to the perceptual field is much stronger than input to the VWM field. Once activation in the perceptual field goes above 0, strong activation is propagated to both the inhibitory and the VWM fields. The VWM field also projects excitatory activation to the inhibitory field, which projects inhibition to both the perceptual and the VWM fields. The model also contains a response layer consisting of two nodes: one that receives summed excitatory input from the perceptual field and is responsible for generating “different” (“Diff”) responses, and a second that receives summed excitatory input from VWM and is responsible for generating “same” responses. The nodes in the response layer have self-excitatory connections and are mutually inhibitory. Note that only above-threshold activation (i.e., activation  $> 0$ ) in the perceptual field or VWM is propagated to the response nodes at test.

(1998) on the basis of studies of cortical neurophysiology. The model consists of an excitatory *perceptual field*, an excitatory *working memory field* (VWM), and a shared *inhibitory field*. As its name suggests, the perceptual field is the main target of afferent input to the network. VWM also receives direct stimulus input, but its primary excitatory input comes from the perceptual field. Both the perceptual field and VWM provide excitatory input to and receive broad inhibitory feedback from the inhibitory field. Additionally, nearby neurons within both the perceptual and the working memory fields interact via local excitatory connections. This pattern of excitatory and inhibitory connectivity gives rise to a “Mexican hat” form of interaction common in neural models of cortical function (Durstewitz, Seamans, & Sejnowski, 2000). With the right balance of excitation and inhibition, multiple peaks of activation can be sustained in the absence of input. (Videos S1 and S2 in the supporting information available on-line show the three-layer model operating, respectively, in a *self-stabilized* mode, in which peaks of activation form in response to input but die out when

input is removed, and in a *self-sustained* mode, in which peaks of activation are sustained in the absence of input; see p. XXX.) Thus, this form of interaction represents a plausible neural basis for the sustained activation proposed to underlie working memory (Compte et al., 2000; Fuster & Alexander, 1971).

Finally, to capture performance in change-detection tasks, we have added a response layer containing two nodes: a *different* node, which receives summed excitatory activation from the perceptual field, and a *same* node, which receives summed excitatory activation from VWM (see Fig. 2b). The nodes are equipped with self-excitatory connections and are mutually inhibitory, competing for control of response output when a “go” signal arrives (following the presentation of the test display).

Visual comparison is made possible in this architecture through excitatory and inhibitory interactions among the model’s layers. Consider the simulations shown in Figure 3, which capture performance in the one-shot variant of the change-detection task (Fig. 1). We focus on this variant of the task because of its relative simplicity, which minimizes the impact of factors



**Fig. 3.** Simulation showing the generation of “same” and “different” responses in the dynamic neural field model of visual working memory (VWM) and change detection. For simplicity, only the two excitatory layers of the model are shown here, although the inhibitory layer plays a critical role in the formation and maintenance of peaks and in the model’s ability to detect changes at test. Following the presentation of a sample input representing two similar colors and one distinctive color (a), three peaks of activation form very quickly in the perceptual field and more slowly in VWM (because input to the perceptual field is stronger). Once activation goes above threshold (0) in the perceptual field, strong activation is transmitted to the inhibitory and VWM layers, and three above-threshold peaks are established in VWM. When the input is removed during the delay interval (b), the peaks die out in the perceptual field, but are sustained in VWM. Inhibitory feedback from VWM to the perceptual field via the inhibitory layer suppresses the firing of neurons in the perceptual field that code for the same features being held in VWM. When a close (c) or far (e) item is probed at test and the input matches one of the remembered features, inhibitory feedback to the perceptual field prevents a new peak from forming. Thus, input to the response nodes comes exclusively from the VWM field, and a “same” (S) response is generated. In contrast, when one of the close items is changed to a new value at test (d), input comes in at a relatively uninhibited region of the perceptual field, allowing a new peak to be established and activation to flow to the “different” (D) node, which wins the competition when a sufficiently strong peak is present in the perceptual field at test. However, when the far item is changed by an identical amount at test (f), input again comes in at a relatively uninhibited region of the perceptual field, but activation is unable to go above threshold, and the model incorrectly responds “same.” Strong laterally inhibitory interactions between close peaks in VWM result in the inhibitory projection to the perceptual field being stronger for far than for close items (compare inhibition in the perceptual field during the delay interval for close vs. far items). The higher level of inhibition makes it more difficult to detect changes to far items.

contributing to failures of change detection when real-world scenes are used as stimuli (see, e.g., Hollingworth, 2003; Hollingworth et al., 2001).

The simulations in Figure 3 demonstrate how “same” and “different” responses arise in the model. Each column shows the pattern of activation in the excitatory layers of the model at a given point in time during a trial in the change-detection task.

Note that, for simplicity, the inhibitory layer is not shown. At the beginning of the trial (Fig. 3a), the model is presented with three inputs: two nearby inputs representing very similar, or “close,” colors and a third input representing a distinct, or “far,” color. When input is turned on, strong activation is applied to the perceptual field, and weaker activation is applied to VWM. Once activation in the perceptual field reaches a given threshold

(conventionally set to be 0), locally excitatory interactions are engaged, and strong activation begins to flow to the inhibitory and VWM fields. Local excitation and reciprocal interactions between the perceptual and inhibitory fields allow three peaks of activation to form in the perceptual field. Shortly thereafter, three peaks of activation also begin to form in VWM. When the input is turned off (Fig. 3b), the peaks quickly die out in the perceptual field. In contrast, the peaks that have now formed in the VWM layer are sustained.

At this point, the only activation entering the perceptual field is inhibitory feedback from the inhibitory layer (not shown in Fig. 3). This input, which is driven by the peaks present in VWM, suppresses the resting level of neurons tuned to the feature values being maintained in VWM (for evidence of perceptual suppression in the context of verbal working memory, see Woodward et al., 2006). When a test item that matches one of the colors in memory is presented (see Fig. 3c for a close item and Fig. 3e for the far item), activation remains below threshold (i.e., below 0) in the perceptual field because the neurons coding for that color are strongly inhibited (see Video S3 in the on-line supporting information). In this case, input to the response layer comes from VWM, allowing the *same* node to win the competition. In contrast, when the close color input is changed at test—to a value  $30^\circ$  away in color space (Fig. 3d)—input enters the perceptual field at a relatively uninhibited site. Consequently, an above-threshold (i.e., above 0) peak forms in the perceptual field at test (see Video S4 in the on-line supporting information), providing input to the *different* node that is strong enough to generate a “different” response.

In summary, responding “different” relies on a distinctive response to novel input, whereas “same” represents the default response of the model (in keeping with the findings of Hyun et al., in press). This aspect of the model is also consistent with classic studies exploring same/different perceptual decisions, which have suggested that “same” and “different” judgments rely on distinct processes (see the review by Farell, 1985). Our model shows how these distinct processes may emerge from the functioning of a single, integrated dynamic system.

### A NOVEL BEHAVIORAL PREDICTION

In addition to providing a plausible neural basis for the processes involved in change detection, our model makes a novel prediction—that the detection of changes in a visual array will be enhanced when metrically similar features are maintained in VWM. To see how this prediction arises, compare the simulation depicted in Figure 3d with that in Figure 3f. Recall that in Figure 3d, we changed one of the close colors by  $30^\circ$  in color space at test, and the model responded “different.” Figure 3f illustrates what happens if we change the far color by an identical amount: A peak fails to build in the perceptual field, and the model erroneously responds “same” (see Video S5 in the on-line supporting information). Why does this occur?

When the peaks in VWM are near one another, they interact in a strongly inhibitory fashion, so that they are somewhat sharper and of somewhat lower amplitude than peaks that are farther apart. As a result, they project weaker excitation to the inhibitory field, which, in turn, projects weaker and narrower inhibition to the perceptual field (see the perceptual field in Fig. 3b). Weaker inhibition makes it easier to build a peak in the perceptual field when a close (Fig. 3d), rather than a far (Fig. 3f), color changes at test.

The similarity-based enhancement predicted by the model runs counter to the predictions of several prominent theories of working memory, which hold that items in working memory are stored independently and do not interact (O’Reilly, Mozer, Munakata, & Miyake, 1999; Raffone & Wolters, 2001). Thus, if confirmed, this novel prediction will have important consequences for neural models of working memory. To explore this prediction further, we conducted a simulation experiment in which the model was run through a standard change-detection experiment.

## MODEL SIMULATIONS

### Method

The model consisted of the architecture shown in Figure 2b (for further details, see Table S1 and Model Architecture, Equations, and Supplementary Simulations in the on-line supporting information).

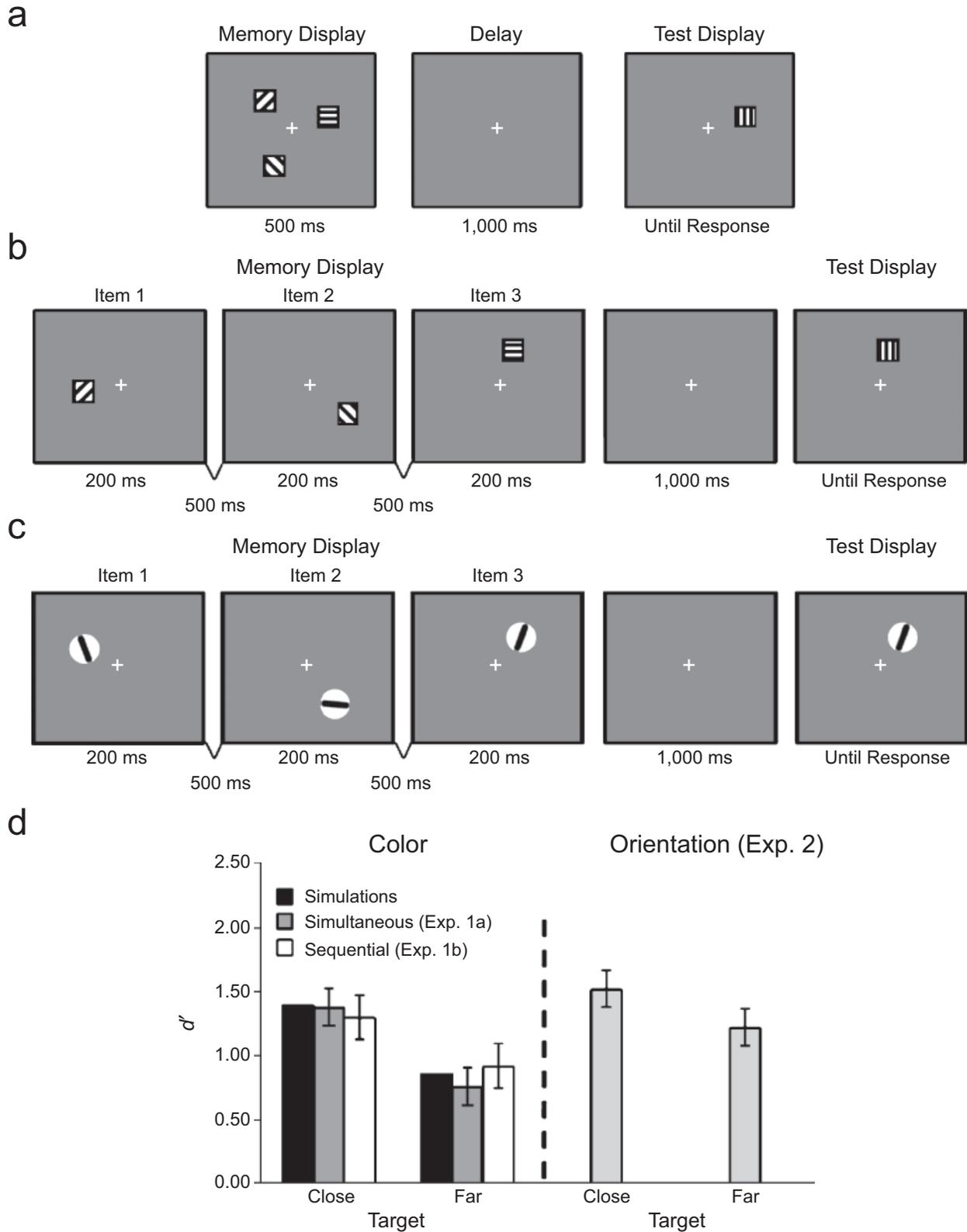
Simulations were conducted in Matlab 7.4 (Mathworks, Inc., <http://www.mathworks.com>). On each trial, the model was presented with three inputs: two that were near each other ( $20^\circ$ ,  $30^\circ$ , or  $40^\circ$  apart, where 1 unit in the model =  $1^\circ$  in color space) and one that was at least  $150^\circ$  away from the nearest close input. At test, a single input was applied at either the same field location as one of the original inputs or a different location  $30^\circ$  away from one of the original inputs, as in the simulations depicted in Figure 3. Thus, memory for either a close target or a far target was probed. For each level of close separation, the model completed 400 trials, 100 change and 100 no-change trials for each target type. Thus, the model completed a total of 1,200 trials.

### Results and Discussion

For each trial, we determined which response node was above threshold at test. We then calculated the signal detection sensitivity measure  $d'$  using the obtained rates of hits and false alarms, as with human subjects. The simulations demonstrated a clear advantage when one of the close items, rather than the far item, was probed at test (see Fig. 4d).

## EXPERIMENTS 1A AND 1B

To test for similarity-based enhancement in human subjects, we conducted a series of change-detection experiments probing



**Fig. 4.** Illustration of the trial sequences in the experiments and results from the experiments and model simulations. Experiment 1a (a) used a standard one-shot change-detection task. A test display of three colors was followed by a delay and then a test display of a single color. Experiment 1b (b) tested change detection when the color stimuli to be remembered were presented sequentially, rather than simultaneously. In these illustrations, different fill patterns represent different solid colors. Experiment 2 (c) followed the procedure for Experiment 1b, but using orientation stimuli. The graph (d) shows performance ( $d'$ ) for close and far targets separately for each experiment and the simulations. Note that for the orientation experiment, results are shown only for trials on which the close-orientation separation was  $20^\circ$ . Error bars represent 95% confidence intervals.

memory for color. In Experiment 1a, memory items were presented simultaneously, as in the standard one-shot change-detection task. Given this presentation format, any observed enhancement might reflect differential perceptual encoding of close versus far colors, rather than interactions among WM representations (see the discussion in Lin & Luck, in press). To rule out this possibility, we presented items sequentially in Experiment 1b.

## Method

### Participants

Ten University of Iowa undergraduates (9 women, 1 man) volunteered to participate in Experiment 1a, and 12 (7 women, 5 men) volunteered to participate in Experiment 1b. Participants received class credit or monetary compensation for their participation. They reported normal or corrected-to-normal visual acuity and normal color vision.

### Stimuli

Stimulus presentation and response recording were controlled by a Macintosh G4 computer running Matlab 5.2 using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Stimuli were presented against a gray background ( $28.73 \text{ cd/m}^2$ ) at a viewing distance of 57 cm. They consisted of small colored squares subtending either  $1.7^\circ \times 1.7^\circ$  (Experiment 1a) or  $2.0^\circ \times 2.0^\circ$  (Experiment 1b). Individual colors were selected from a set of 180 colors equally distributed in CIELAB 1976 color space, a perceptually uniform color space and color-appearance model developed by the Commission Internationale de l'Éclairage.

The memory displays consisted of three items presented at least  $90^\circ$  apart on the circumference of an imaginary circle that was centered at fixation and had a radius of either 4.25 cm (Experiment 1a) or 7.5 cm (Experiment 1b). In each case, two of the items were close in color space ( $20^\circ$ ,  $30^\circ$ , or  $40^\circ$  apart), whereas the color of the third item was always at least  $150^\circ$  away from the nearest close color. The test display contained a single item appearing at one of the spatial locations previously occupied by an item in the memory display (see Fig. 4a). On change trials, the color of the test input and the memory item always differed by  $30^\circ$  in color space. Close and far items were tested equally often.

### Procedure

In Experiment 1a (see Fig. 4a), each trial began with a fixation cross presented at the center of the screen for 500 ms. Next, the memory display was presented for 500 ms, followed by a 1,000-ms delay interval and then a test display, which remained visible until a response was generated. Participants were instructed to make an unspeeded response when the test display appeared, indicating whether the color of the test item was the same as or different from the item appearing at that location originally.

Participants completed a practice block of 24 trials and a total of 240 experimental trials: 40 trials with close targets and 40 trials with far targets at each level of close-color separation (probability of change = .5).

In Experiment 1b (see Fig. 4b), each trial also began with a 500-ms fixation cross, and then the memory items were presented sequentially for 200 ms each, separated by 500-ms blank intervals. Close and far targets appeared with equal likelihood in each probe position (1, 2, or 3) and were probed an equal number of times at each position. Participants completed 48 trials (24 change and 24 no-change trials) for each combination of target type (close, far), probe position (1, 2, 3), and close-color separation ( $20^\circ$ ,  $30^\circ$ ,  $40^\circ$ ), for a total of 864 trials.

To prevent verbal recoding of the memory-display colors, we instructed participants to repeat three randomly generated digits (e.g., “6, 4, 9”) out loud at a regular pace throughout each trial of each experiment.

## Results and Discussion

An alpha level of .05 was used as the criterion for statistical significance, and the signal detection measure  $d'$  was the primary dependent measure.

### Experiment 1a

An analysis of variance (ANOVA) with target type (close, far) and close-color separation ( $20^\circ$ ,  $30^\circ$ ,  $40^\circ$ ) as within-subjects factors revealed a significant main effect of target type,  $F(1, 9) = 29.82$ ,  $p < .001$ . No other effects reached statistical significance, all  $F$ s  $< 1$ . As predicted, change-detection performance was enhanced when a close, rather than a far, color was probed (see Fig. 4d).

### Experiment 1b

An ANOVA with target type, probe position, and close-color separation as within-subjects factors revealed a main effect of probe position,  $F(2, 22) = 14.60$ ,  $p < .001$ ; change detection was best for the most recent item (mean  $d' = 0.88$ , 1.0, and 1.47 for Positions 1, 2, and 3, respectively). There was also a main effect of close-color separation,  $F(2, 22) = 17.99$ ,  $p < .001$ ; change detection was most accurate when the close colors were separated by  $20^\circ$  (mean  $d' = 1.30$ , 1.12, and 0.92 for separations of  $20^\circ$ ,  $30^\circ$ , and  $40^\circ$ , respectively). Most critically, there was a main effect of target type,  $F(1, 11) = 24.01$ ,  $p < .001$ . As in Experiment 1a, change-detection performance was enhanced when a close, rather than a far, color was probed (see Fig. 4d). These results replicate those of Experiment 1a, confirming that the enhancement effect does not reflect differential perceptual encoding of close versus far colors.

## EXPERIMENT 2

Results from Experiment 1 are consistent with the predictions of the dynamic field model proposed here. In Experiment 2,

we examined whether the enhancement effect is a general property of the neural mechanisms that underlie VWM by probing whether this effect generalizes to working memory for orientation.

## Method

### *Participants*

Fifteen University of Iowa undergraduates (10 women, 5 men) volunteered to participate.

### *Stimuli and Procedure*

Stimulus presentation, response recording, and the procedure were the same as in Experiment 1b. The memory items consisted of three thin, black, rounded rectangles, each of which spanned the interior of a small, light-gray circle (2.0° in diameter; see Fig. 4c). On each trial, two of the memory items had similar orientations (20°, 30°, or 40° apart), whereas the orientation of the third item differed by at least 70° from the nearest close orientation. When a change occurred at test, the orientation of the test item was rotated by 30°.

## Results and Discussion

An ANOVA with target type, probe position, and close-orientation separation as within-subjects factors revealed a main effect of probe position,  $F(2, 28) = 30.24, p < .001$ . Again, change-detection performance was best for the most recently presented item (mean  $d' = 1.01, 1.12, \text{ and } 2.00$  for Positions 1, 2, and 3, respectively). In addition, there was a main effect of close-orientation separation (20°, 30°, 40°),  $F(2, 28) = 3.40, p < .05$ , which was subsumed by a significant Target Type  $\times$  Separation interaction,  $F(2, 28) = 4.06, p < .03$ . Tests of simple effects revealed significantly better performance for close than for far targets when the close-orientation separation was 20° (see Fig. 4d),  $F(1, 2) = 7.67, p < .02$ , but not when this separation was 30° or 40°, all  $F_s < 1$ . This finding supports the proposal that similarity-based enhancement is a general property of working memory for metric feature dimensions.

## GENERAL DISCUSSION

We have described a neurally grounded model that incorporates multi-item encoding and maintenance in VWM, as well as the processes underlying visual comparison. The model achieves sustained activation via locally excitatory and laterally inhibitory interactions among neurons, and interactions among the model's layers give rise to an emergent form of comparison that drives decisions about detected change. Specifically, peaks of activation in VWM pass activation to an inhibitory field, and the inhibitory field in turn provides inhibitory feedback to a perceptual field that is responsible for the detection of novel inputs. When new inputs match the remembered features, a peak fails to

build in the perceptual field, and a “same” response is triggered. In contrast, when a change occurs at test, a new peak is formed in the perceptual field, and this peak triggers a “different” response. Thus, detecting difference depends on an active change signal that may serve to direct attention or the eyes to the location of potentially interesting changes in the environment (Hyun et al., in press).

The present study also revealed a direct behavioral consequence of the metric-dependent neural processes supporting maintenance and comparison in the DFT model proposed here: enhanced change detection when metrically similar features are remembered. Recall that the transition from a “same” to a “different” response, which arises when sensory inputs are sufficiently different from memory representations, is mediated via inhibition of the feature values associated with each memory item. The same inhibitory source also produces lower-amplitude peaks when two close items, rather than two far items, are held in working memory simultaneously. This weakening reduces the strength of inhibition passed to the perceptual field, producing the predicted enhancement effect. Thus, the metric dependence of change detection is not a side effect or generic property of systems with lateral inhibition, but is linked to the core mechanism supporting change detection in the model. The interitem metric interactions proposed to underlie this effect run counter to several prominent theories according to which items in working memory are represented independently and do not interact (see, e.g., O'Reilly et al., 1999; Raffone & Wolters, 2001). Thus, our findings place strong constraints on models of the nature of VWM and its underlying mechanisms.

The model described here suggests a candidate neural mechanism for the explicit comparison process proposed to underlie change detection (Mitroff et al., 2004). Additionally, because our model makes a “same” or “different” decision on each trial, it can offer insights into the origin of failures to detect change, so-called change blindness. For instance, errors can arise in our model if a change is too small to be detected, or if the relevant items are not successfully encoded and maintained in working memory (for related discussion, see Hollingworth, 2003; Mitroff et al., 2004).

One final question concerns how our functional neural model maps onto evidence of neural localization from neurophysiological and functional imaging studies. The three-layered architecture we propose is presumed to operate within the laminar layers of a given cortical area—for instance, within V4 or inferior temporal cortex. It is also possible, however, to achieve the same functionality using a four-layer architecture in which each excitatory layer projects to a local inhibitory population in addition to the inhibitory population of the other excitatory field (see, e.g., Edin, Macoveanu, Olesen, Tegner, & Klingberg, 2007). This four-layered architecture would be consistent with the proposal that the perceptual field resides in posterior cortex and the working memory field resides in another area, such as the prefrontal cortex. Future work will be required to assess

which architecture is more appropriate. For the present, however, the DFT provides a critical bridge between neurophysiological measures and the details of human performance in change-detection tasks.

**Acknowledgments**—This research was made possible by National Institute of Mental Health Grant 2R01-MH062480 and National Science Foundation Grant HSD0527698, awarded to J.P.S., and by National Institute of Mental Health Grant R01-MH076226 awarded to S.J.L.

## REFERENCES

- Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, *27*, 77–87.
- Amit, D.J., Bernacchia, A., & Yakovlev, V. (2003). Multiple-object working memory: A model for behavioral performance. *Cerebral Cortex*, *13*, 435–443.
- Beck, D.M., Rees, G., Frith, C.D., & Lavie, N. (2001). Neural correlates of change detection and change blindness. *Nature Neuroscience*, *4*, 645–650.
- Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.
- Compte, A., Brunel, N., Goldman-Rakic, P.S., & Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, *10*, 910–923.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.
- Douglas, R., & Martin, K. (1998). Neocortex. In G.M. Shepherd (Ed.), *The synaptic organization of the brain* (4th ed., pp. 459–510). New York: Oxford University Press.
- Durstewitz, D., Seamans, J.K., & Sejnowski, T.J. (2000). Neurocomputational models of working memory. *Nature*, *3*, 1184–1191.
- Edin, F., Macoveanu, J., Olesen, P., Tegner, J., & Klingberg, T. (2007). Stronger synaptic connectivity as a mechanism behind development of working memory-related brain activity during childhood. *Journal of Cognitive Neuroscience*, *19*, 750–760.
- Farell, B. (1985). “Same”-“different” judgements: A review of current controversies in perceptual comparison. *Psychological Bulletin*, *98*, 419–456.
- Fuster, J.M., & Alexander, G. (1971). Neuron activity related to short-term memory. *Science*, *173*, 652–654.
- Gegenfurtner, K.R., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 845–866.
- Grossberg, S. (1980). Biological competition: Decision rules, pattern formation and oscillations. *Proceedings of the National Academy of Sciences, USA*, *77*, 2338–2342.
- Hollingworth, A. (2003). Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 388–403.
- Hollingworth, A., Shrock, G., & Henderson, J.M. (2001). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory & Cognition*, *29*, 296–304.
- Hyun, J.-S., Woodman, G.F., Vogel, E.K., Hollingworth, A., & Luck, S.J. (in press). The comparison of visual working memory representations with perceptual inputs. *Journal of Experimental Psychology: Human Perception and Performance*.
- Jancke, D., Erilhagen, W., Dinse, H.R., Akhavan, A.C., Giese, M.A., Steinhage, A., & Schöner, G. (1999). Parametric population representation of retinal location: Neuronal interaction dynamics in cat primary visual cortex. *Journal of Neuroscience*, *19*, 9016–9028.
- Jolicoeur, P., & Dell’Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology*, *36*, 138–202.
- Lin, P.-H., & Luck, S.J. (in press). The influence of similarity on visual working memory representations. *Visual Cognition*.
- Luck, S.J. (in press). Visual short-term memory. In S.J. Luck & A. Hollingworth (Eds.), *Visual memory*. New York: Oxford University Press.
- Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.
- Mitroff, S.R., Simons, D.J., & Levin, D.T. (2004). Nothing compares two views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics*, *66*, 1268–1281.
- O’Reilly, R.C., Mozer, M., Munakata, Y., & Miyake, A. (1999). Discrete representations in working memory: A hypothesis and computational investigations. In *The Second International Conference on Cognitive Science* (pp. 183–188). Tokyo: Japanese Cognitive Science Society.
- Pashler, H. (1988). Familiarity and the detection of change in visual displays. *Perception & Psychophysics*, *44*, 369–378.
- Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Pessoa, L., Gutierrez, E., Bandettini, P.B., & Ungerleider, L.G. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron*, *35*, 975–987.
- Pessoa, L., & Ungerleider, L.G. (2004). Neural correlates of change detection and change blindness in a working memory task. *Cerebral Cortex*, *14*, 511–520.
- Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience*, *13*, 766–785.
- Rensink, R.A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277.
- Schmidt, B.K., Vogel, E.K., Woodman, G.F., & Luck, S.J. (2002). Voluntary and automatic attentional control of visual working memory. *Perception & Psychophysics*, *64*, 754–763.
- Simmering, V.R., Schutte, A.R., & Spencer, J.P. (2008). Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. *Brain Research*, *1202*, 68–86.
- Simons, D.J., & Rensink, R.A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*, 16–20.
- Spencer, J.P., Simmering, V.R., Schutte, A.R., & Schöner, G. (2007). What does theoretical neuroscience have to offer the study of behavioral development? Insights from a dynamic field theory of spatial cognition. In J.M. Plumert & J.P. Spencer (Eds.), *The emerging spatial mind* (pp. 320–361). New York: Oxford University Press.
- Todd, J.J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*, 751–754.

- Vogel, E.K., & Machizawa, M.G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.
- 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*, 1436–1451.
- Wilson, H.R., & Cowan, J.D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*, 1–24.
- Woodward, T.S., Cairo, T.A., Ruff, C.C., Takane, Y., Hunter, M.A., & Ngan, E.T.C. (2006). Functional connectivity reveals load dependent neural systems underlying encoding and maintenance in verbal working memory. *Neuroscience*, *139*, 317–325.
- Xu, Y., & Chun, M.M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95.
- Zhang, W., & Luck, S.J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.

(RECEIVED 12/26/07; REVISION ACCEPTED 9/27/08)

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the on-line version of this article:

**Model Architecture, Equations, and Supplementary Simulations**

**Table S1**

**Video S1**

**Video S2**

**Video S3**

**Video S4**

**Video S5**

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.